Tesi di Dottorato

Università degli Studi di Napoli "Federico II"

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

Dottorato di Ricerca in Ingegneria Elettronica e delle Telecomunicazioni

Change Detection for Multi-Polarization SAR

Vincenzo Carotenuto

Il Coordinatore del Corso di Dottorato	Il Tutore
Ch.mo Prof. Niccoló Rinaldi	Ch.mo Prof. Antonio De Maio

XXVII CICLO

Contents

Introduction

A topic of great interest in the Remote Sensing, Signal Processing, and Synthetic Aperture Radar (SAR) communities is change detection. This is the ability to identify temporal changes within a given scene starting from a pair of co-registered SAR images representing an area of interest [1–3]. Incoherent and coherent change detection are the two main approaches that have been proposed in the open literature to process the image pair. The former attempts to detect changes in the mean power level of a given scene exploiting only the intensity information from the available images (thus neglecting phase information [4–6]): differencing and rationing are well-known techniques in this context [7]. The latter jointly uses both amplitude and phase from the reference and the test data to detect possible changes in the region of interest. In [4,5], a thorough comparison between incoherent and coherent change detection strategies, including the Maximum Likelihood Estimate (MLE) of the SAR coherence parameter, is performed based on high resolution (0.3 m×0.3 m) SAR images. In [7], several techniques for change detection have been presented and compared based on their probability of error and on results obtained using repeat-pass ERS-1 SAR data. In [6, 8–10], the multi-polarization signal model for the SAR change detection problem is laid down, the detection problem is formulated as a binary hypothesis test, and a decision rule based on the Generalized Likelihood Ratio Test (GLRT) is developed. Moreover, the performance analysis [9] of the GLRT is given in the form of Receiver Operating Characteristics (ROC), namely detection Probability (P_d) versus false alarm Probability (P_{fa}), quantifying the benefits of the multi-polarization information in SAR change detection. A complementary approach to the GLRT is considered in [11] based on the use of perturbation filters and a separated treatment between polarimetry and amplitude.
A detection scheme based on canonical correlations analysis is applied

to scalar EMISAR data in [12, 13], whereas, in [14], a mutual information based framework is developed to address coherent similarities between multichannel SAR images. Starting from the multi-polarization data model developed in [8] and [9], a new and systematic framework for change detection based on the theory of invariance in hypothesis testing problems [15, 16] is proposed. This is a viable mean to force some desired properties to a decision statistic at the design stage and has already been successfully applied in some different radar detection problems [17–19]. Otherwise stated, the principle of invariance allows to focus on decision rules which exhibit some natural symmetries implying important practical properties such as the Constant False Alarm Rate (CFAR) behavior. Besides, the use of invariance leads to a data reduction because all invariant tests can be expressed in terms of a statistic, called maximal invariant, which organizes the original data into equivalence classes. Also the parameter space is usually compressed after reduction by invariance and the dependence on the original set of parameters become embodied into a maximal invariant in the parameter space (induced maximal invariant).

Therefore, the thesis is organized as follows:

- In Chapter 1, with reference to the multi-polarization SAR change detection problem, we first determine a maximal invariant statistic in terms of the eigenvalues of a data dependent matrix constructed using both test and reference images. Then, we design the Most Powerful Invariant (MPI) receiver, as the Neyman-Pearson detector computed from the maximal invariant statistic, and show that no Uniformly MPI (UMPI) test exists, namely the optimum invariant detector is not implementable. Hence, we focus on suboptimum invariant receivers based on different functions of the maximal invariant such as arithmetic mean, harmonic mean, maximum and minimum entries. Remarkably, the GLRT in [8], [9], belongs to the proposed class of receivers. At the analysis stage, we assess the performance of the considered invariant decision rules in correspondence of a two-channel polarization diversity providing detection probability contours, for a given false alarm level, versus the induced maximal invariant which turns out bi-dimensional. We also conduct a performance comparisons with the benchmark MPI test showing that some analyzed receivers can outperform the GLRT and achieve a detection probability close to that en-

sured by the clairvoyant MPI structure. Besides, we consider ROC curves and assess the impact of the polarimetric information on the detection performance. Finally, to validate the behavior of the considered tests on real SAR images, we use a high resolution change detection dataset, available from Air Force Research Laboratory (AFRL) [20, 21] and collected from an X-band SAR. The conducted analysis highlights the detectors capability to operate successfully in real environments; some of these new detectors have the interesting feature to directly discriminate between car arrivals and departures without any additional Radar Cross Section (RCS) comparison.

- In Chapter 2, we exploit the block-diagonal structure for the polarimetric covariance matrix introduced in [22] and devise a decision rule based on the GLRT criterion. Remarkably, it ensures the CFAR property with respect to the unknown polarimetric covariance provided that it complies with a certain design structure. The benchmark clairvoyant optimum detector and the GLRT derived in [9] (without the exploitation of the special covariance structure forced by polarization diversity) are used as benchmark at the analysis stage to assess the performance of the new structured GLRT. The analysis, conducted both on simulated as well as on real high resolution SAR data, shows the effectiveness of the structured GLRT and its superiority over the classic unstructured GLRT is some scenarios of practical interest.

- In Chapter 3, starting from the framework proposed in Chapter 1, we introduce the capability to account for a possible scale mismatch factor. The new approach is able to produce scale-invariant decision rules, providing advantages in terms of robustness to intensity mismatches and/or miscalibrations and false alarm rejection with respect to those considered in Chapter 1. This is an important property as images over the same scene can exhibit different intensity scales due to different observation angles and propagation properties. As a matter of fact, there are situations in Earth Observation where a change detector without scale-invariance is doomed to fail. A classic example is the sea where different sea states can trigger a detector without scale invariance. Another example may be a ground surface that shows very different moisture

conditions (due for instance to rain). The mentioned effects can often lead to false alarms in a change detection structure that is not designed to be robust with respect to such scale variations. We also compute the GLRT detector and prove that it belongs to the class of scale-invariant decision rules. Interestingly, with reference to two-polarimetric channels, it is tantamount to comparing the condition number of a data-dependent matrix with a suitable threshold. From a technical (detection theory) point of view, the derivations of the GLRT receiver for the considered problem are, to the best of authors knowledge, a new result together with the analytic expressions of P_{fa} and P_d (developed in Appendix B) and the determination of the maximal invariant statistic. At the analysis stage, we assess the performance of the considered invariant decision rules in correspondence of a two- (and three-) channel polarization diversity providing detection probability contours, for a given false alarm level, versus the induced maximal invariant which turns out one- (bi-) dimensional, ROC curves and the value of the P_{fa} for different scale values. The analysis, conducted both on simulated as well as on real high resolution SAR data, highlights the capability of the proposed detectors to provide scale invariance and, at the same time, ensuring satisfactory detection performances.

At the end of each chapter, some conclusions and possible future tracks are given.

Notation

$\boldsymbol{a}$	column vector;
$a(i)$	i-th element of the column vector $\boldsymbol{a}$;
$\boldsymbol{a}_i$	i-th column vector;
$\boldsymbol{A}$	matrix;
$A(i,k)$	(i,k)-th entry of the matrix $\boldsymbol{A}$;
$(\cdot)^T$	transpose operator;
$(\cdot)^*$	complex conjugate operator (component-wise complex conjugate if the argument is a matrix or a vector);
$(\cdot)^\dagger$	transpose conjugate operator;
$\mathrm{tr}(\cdot)$	trace of the square matrix argument;
$\det(\cdot)$	determinant of the square matrix argument;
$\mathbf{diag}(\cdot)$	vector formed by the diagonal elements of the matrix argument;
$\mathbf{Diag}(\cdot)$	diagonal matrix formed by the components of the vector argument;
$\boldsymbol{I}$	identity matrix;
$\mathbf{0}$	matrix with zero entries;
$\mathbf{1}_N$	$N \times 1$ vector with all the entries equal to one;
$\succ$	generalized matrix inequality: for any Hermitian matrix $\boldsymbol{A}$, $\boldsymbol{A} \succ \mathbf{0}$ means that $\boldsymbol{A}$ is a positive definite matrix;
$\mathcal{H}_N^{++}$	set of $N \times N$ Hermitian positive definite matrices;
$\mathcal{R}^{++}$	set of positive real numbers;
$\mathcal{GL}(N)$	General Linear group of degree N over the field of complex numbers: is the set of $N \times N$ non-singular matrices together with the operation of ordinary matrix multiplication.

Chapter 1

Invariant Rules for Multi-Polarization SAR Change Detection

This chapter deals with coherent (in the sense that both amplitudes and relative phases of the polarimetric returns are used to construct the decision statistic), multi-polarization SAR change detection assuming the availability of reference and test images collected from N multiple polarimetric channels. At the design stage, the change detection problem is formulated as a binary hypothesis testing problem and the principle of invariance is used to come up with decision rules sharing the Constant False Alarm Rate (CFAR) property. The maximal invariant statistic and the maximal invariant in the parameter space are obtained. Hence, the optimum invariant test is devised proving that a Uniformly Most Powerful Invariant (UMPI) detector does not exist. Based on this, the class of sub-optimum invariant receivers, which also includes the Generalized Likelihood Ratio Test (GLRT), is considered. At the analysis stage, the performance of some tests, belonging to the aforementioned class, is assessed and compared with the optimum clairvoyant invariant detector. Finally, detection maps on real high resolution SAR data are computed showing the effectiveness of the considered invariant decision structures. Precisely, in Section 1.1 we deal with the formulation of the multi-polarization SAR change detection problem, whereas, in Section 1.2, we address maximal invariant and induced maximal invariant design. In Section 1.3, we devise the MPI test and show that a UMPI

detector does not exist for the present problem. Thus, we consider the design of sub-optimum invariant decision rules. In Sections 1.4 and 1.5, we assess the performance of the introduced invariant tests respectively on simulated and on real multi-polarization SAR images. Finally, in Section 1.6, we draw conclusions and outline some possible future research tracks.

1.1 Problem Formulation

A multipolarization SAR sensor measures for each pixel of the image under test $N \in \{2, 3\}$ complex returns, collected from different polarimetric channels (for instance HH and VV for $N = 2$; HH, VV, and HV with reference to $N = 3$). The N returns from the same pixel are stacked to form the vector $\boldsymbol{X}(l, m)$, where $l = 1, \ldots, L$ and $m = 1, \ldots, M$ (L and M represent the vertical and horizontal size of the image, respectively). Therefore, the sensor provides a 3-D data stack $\boldsymbol{X}$ of size $M \times L \times N$ which is referred to in the following as a datacube and is illustrated in Figure 1.1.

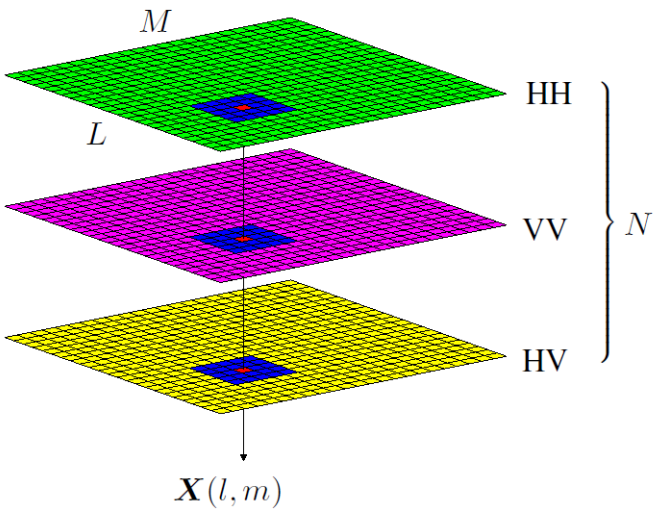


Figure 1.1: Construction of the datacube.

For SAR change detection applications, we suppose that two datacubes $\boldsymbol{X}$ (reference data) and $\boldsymbol{Y}$ (test data) of the same geographic

area are available; they are collected from two different sensor passes and are accurately pixel aligned (co-registrated). We focus on the problem of detecting the presence of possible changes in a rectangular neighborhood $\mathcal{A}$, with size $K = W_1 \times W_2 \geq N$, of a given pixel. To this end, we denote by $\boldsymbol{R}_X$ ($\boldsymbol{R}_Y$) the matrix whose columns are the vectors of the polarimetric returns from the pixels of $\boldsymbol{X}$ ($\boldsymbol{Y}$) which fall in the region $\mathcal{A}$ and $\boldsymbol{S}_X = \boldsymbol{R}_X \boldsymbol{R}_X^\dagger$ ($\boldsymbol{S}_Y = \boldsymbol{R}_Y \boldsymbol{R}_Y^\dagger$).

The matrices $\boldsymbol{R}_X$ and $\boldsymbol{R}_Y$ are modeled as statistically independent random matrices. Moreover, the columns of $\boldsymbol{R}_X$ ($\boldsymbol{R}_Y$) are assumed statistically independent and identically distributed random vectors[1] drawn from a complex circular zero-mean Gaussian distribution with positive definite covariance matrix $\boldsymbol{\Sigma}_X$ ($\boldsymbol{\Sigma}_Y$). From the physical point of view this is tantamount to assuming a fully developed speckle [23, 24]. Under the aforementioned settings, the change detection problem in the region $\mathcal{A}$ can be formulated in terms of the following binary hypothesis test

$$\begin{cases} H_0 : \boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y \\ H_1 : \boldsymbol{\Sigma}_X \neq \boldsymbol{\Sigma}_Y \end{cases} \tag{1.1}$$

where the null hypothesis H_0 of change absence is tested versus the alternative[2] H_1.

Exploiting the Gaussian assumption, we can write the joint probability density function (pdf) of $\boldsymbol{R}_X$ and $\boldsymbol{R}_Y$ as

$$f_{\boldsymbol{R}_X,\boldsymbol{R}_Y}(\boldsymbol{R}_X, \boldsymbol{R}_Y | H_1, \boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y) = \frac{1}{\pi^{2NK} \det(\boldsymbol{\Sigma}_X \boldsymbol{\Sigma}_Y)^K} \exp\left\{-\mathrm{tr}\left(\boldsymbol{\Sigma}_X^{-1} \boldsymbol{S}_X + \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{S}_Y\right)\right\} . \tag{1.2}$$

Using the Fisher-Neyman factorization theorem [25], we can claim that a sufficient statistic for (1.1) is represented by the two sample Grammian matrices $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ which are statistically independent and follow a

[1] As to the spatial independence among the polarimetric returns from different pixels, it may be somewhat limiting especially in the presence of spatial oversampling and after the processing operations required by geocoding. Nevertheless, this assumption is usually met in many multilook SAR signal processing techniques to grant analytic tractability. The goodness of the approximation is a-posteriori assessed when testing the processors on real data.

[2] This testing problem is also well known in statistical literature with reference to real observations [15, Ch. 8].

complex Wishart distribution, i.e. [26]

$$f_{\boldsymbol{S}_X}(\boldsymbol{S}_X|H_1,\boldsymbol{\Sigma}_X) = \frac{c_W}{\det(\boldsymbol{\Sigma}_X)^K}\exp\left\{-\mathrm{tr}\left(\boldsymbol{\Sigma}_X^{-1}\boldsymbol{S}_X\right)\right\}\det(\boldsymbol{S}_X)^{K-N}, \quad \boldsymbol{S}_X \succ \mathbf{0} \tag{1.3}$$

$$f_{\boldsymbol{S}_Y}(\boldsymbol{S}_Y|H_1,\boldsymbol{\Sigma}_Y) = \frac{c_W}{\det(\boldsymbol{\Sigma}_Y)^K}\exp\left\{-\mathrm{tr}\left(\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{S}_Y\right)\right\}\det(\boldsymbol{S}_Y)^{K-N}, \quad \boldsymbol{S}_Y \succ \mathbf{0} \tag{1.4}$$

with c_W a normalization constant. From the sufficient statistic, we can evaluate the optimum Neyman-Pearson (NP) detector as the Likelihood Ratio Test (LRT),

$$\frac{f_{\boldsymbol{S}_X,\boldsymbol{S}_Y}(\boldsymbol{S}_X,\boldsymbol{S}_Y|H_1,\boldsymbol{\Sigma}_X,\boldsymbol{\Sigma}_Y)}{f_{\boldsymbol{S}_X,\boldsymbol{S}_Y}(\boldsymbol{S}_X,\boldsymbol{S}_Y|H_0,\boldsymbol{\Sigma}_X)} = \frac{f_{\boldsymbol{S}_X}(\boldsymbol{S}_X|H_1,\boldsymbol{\Sigma}_X)f_{\boldsymbol{S}_Y}(\boldsymbol{S}_Y|H_1,\boldsymbol{\Sigma}_Y)}{f_{\boldsymbol{S}_X}(\boldsymbol{S}_X|H_0,\boldsymbol{\Sigma}_X)f_{\boldsymbol{S}_Y}(\boldsymbol{S}_Y|H_0,\boldsymbol{\Sigma}_X)} \underset{H_0}{\overset{H_1}{\gtrless}} T_{0,A}\,,$$

where the first equality stems from the statistical independence between $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$; $T_{0,A}$ is the detecton threshold set to ensure a given P_{fa} value. Substituting the pdfs of $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ in the LRT, after standard algebra and statistical equivalences (also absorbing irrelevant data-independent terms in the threshold), it can be recast as

$$\mathrm{tr}\left[\left(\boldsymbol{\Sigma}_X^{-1}-\boldsymbol{\Sigma}_Y^{-1}\right)\boldsymbol{S}_Y\right] \underset{H_0}{\overset{H_1}{\gtrless}} T_0 \tag{1.5}$$

where T_0 is the suitable modification of the original detection threshold. Evidently, test (1.5) is not Uniformly Most Powerful (UMP) and, consequently, it is not practically implementable because it requires the knowledge (clairvoyant detector) of both $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ which, in realistic applications, are usually unknown.

1.2 Data Reduction and Invariance Issues

Both hypotheses under test are composite or, in other words, H_0 and H_1 are equivalent to a partition of the parameter space Θ into the two disjoint sets

$$\begin{aligned}\Theta_0 &= \left\{\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y,\ (\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y) \in \mathcal{H}_N^{++} \times \mathcal{H}_N^{++}\right\} \\ \Theta_1 &= \left\{\boldsymbol{\Sigma}_X \neq \boldsymbol{\Sigma}_Y,\ (\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y) \in \mathcal{H}_N^{++} \times \mathcal{H}_N^{++}\right\}\end{aligned} \tag{1.6}$$

This formulation emphasizes that the individual values of the nuisance parameters are irrelevant: one must only discern to which hypothesis they pertain, namely whether the covariances are equal or not. This suggests that we can cluster the data considering transformations that leave unaltered:

a. the two composite hypotheses, namely the partition of the parameter space;

b. the families of distributions under the two hypotheses.

This goal can be achieved through the *Principle of Invariance* [16]. According to such a principle, we look for transformations that preserve the formal structure of the hypothesis testing problem and, then, for decision rules invariant to them. Such a principle also acts as a data reduction technique leading to a reduced observation space of significantly lower dimensionality than the original one.

It is not difficult to prove that our testing problem is invariant under the group of transformations G acting on the sufficient statistic as:

$$G = \left\{g\ :\ \boldsymbol{S}_X \to \boldsymbol{B}\boldsymbol{S}_X\boldsymbol{B}^\dagger\,, \quad \boldsymbol{S}_Y \to \boldsymbol{B}\boldsymbol{S}_Y\boldsymbol{B}^\dagger\,, \quad \boldsymbol{B} \in \mathcal{GL}(N)\right\}. \tag{1.7}$$

In fact, the families of distributions are preserved because if $\boldsymbol{S}_X$ ($\boldsymbol{S}_Y$) is Wishart distributed also $\boldsymbol{B}\boldsymbol{S}_X\boldsymbol{B}^\dagger$ ($\boldsymbol{B}\boldsymbol{S}_Y\boldsymbol{B}^\dagger$) is Wishart with the same scalar parameters and matrix parameter $\boldsymbol{B}\boldsymbol{\Sigma}_X\boldsymbol{B}^\dagger$ ($\boldsymbol{B}\boldsymbol{\Sigma}_Y\boldsymbol{B}^\dagger$), where $\boldsymbol{B} \in \mathcal{GL}(N)$. Moreover, the original partition of the parameter space is left unaltered since if $\boldsymbol{\Sigma}_X \neq \boldsymbol{\Sigma}_Y$ then $\boldsymbol{B}\boldsymbol{\Sigma}_X\boldsymbol{B}^\dagger \neq \boldsymbol{B}\boldsymbol{\Sigma}_Y\boldsymbol{B}^\dagger$ and if $\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y$ then $\boldsymbol{B}\boldsymbol{\Sigma}_X\boldsymbol{B}^\dagger = \boldsymbol{B}\boldsymbol{\Sigma}_Y\boldsymbol{B}^\dagger$.

1.2.1 Maximal Invariant Design

The invariance property induces a partition of the data space into orbits (or equivalence classes) where, over each orbit, every point is related to every other through a transformation which is a member of the group G. Any statistic that identifies different orbits in a one-to-one way significantly reduces the total amount of data necessary for

solving the hypothesis testing problem and constitutes the compressed data set to be used in the design of any invariant detector. These kind of statistics are called maximal invariants since they are constant over each orbit (invariance) while assuming different values on different orbits (maximality).

Formally, a statistic $\mathbf{T}(\boldsymbol{S}_X, \boldsymbol{S}_Y)$ is said to be a maximal invariant with respect to the group of transformations G if and only if

- **Invariance**: $\mathbf{T}(\boldsymbol{S}_X, \boldsymbol{S}_Y) = \mathbf{T}[g(\boldsymbol{S}_X, \boldsymbol{S}_Y)]$, $\forall g \in G$.
- **Maximality**: $\mathbf{T}(\boldsymbol{S}_{X_1}, \boldsymbol{S}_{Y_1}) = \mathbf{T}(\boldsymbol{S}_{X_2}, \boldsymbol{S}_{Y_2})$ implies that $\exists\, g \in G$ such that $(\boldsymbol{S}_{X_2}, \boldsymbol{S}_{Y_2}) = g(\boldsymbol{S}_{X_1}, \boldsymbol{S}_{Y_1})$.

Notice that there are many maximal invariant statistics, but they are equivalent in that yield statistically equivalent detectors. Moreover, all invariant tests can be expressed as a function of the maximal invariant statistic [15, 27], which for the problem of interest is provided by the following

Proposition 1: A maximal invariant statistic for problem (1.1) with respect to the group of transformations (1.7) is the N-dimensional vector of the eigenvalues $\lambda_1, \ldots, \lambda_N$ of

$$\boldsymbol{S}_X \boldsymbol{S}_Y^{-1}\,. \tag{1.8}$$

Proof. The invariance of $\lambda_1, \ldots, \lambda_N$ follows because the eigenvalues of $\boldsymbol{B}\boldsymbol{S}_X\boldsymbol{B}^\dagger(\boldsymbol{B}\boldsymbol{S}_Y\boldsymbol{B}^\dagger)^{-1} = \boldsymbol{B}\boldsymbol{S}_X\boldsymbol{S}_Y^{-1}\boldsymbol{B}^{-1}$ are the same as those of $\boldsymbol{S}_X\boldsymbol{S}_Y^{-1}$. As to the maximality, we have to prove that if $\boldsymbol{S}_{X,1}\boldsymbol{S}_{Y,1}^{-1}$ and $\boldsymbol{S}_{X,2}\boldsymbol{S}_{Y,2}^{-1}$ share the same eigenvalues then there exists a non-singular matrix $\boldsymbol{B}$ such that $\boldsymbol{S}_{X,1} = \boldsymbol{B}\boldsymbol{S}_{X,2}\boldsymbol{B}^\dagger$ and $\boldsymbol{S}_{Y,1} = \boldsymbol{B}\boldsymbol{S}_{Y,2}\boldsymbol{B}^\dagger$. To this end, denoting by $\bar{\boldsymbol{\Lambda}} = \mathbf{Diag}\left([\lambda_1, \ldots, \lambda_N]^T\right)$ and exploiting Theorem [28, Corol. 4.6.12, p. 250], we can claim that there exist two non-singular matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ such that

$$\begin{aligned} \boldsymbol{B}_1\boldsymbol{S}_{X,1}\boldsymbol{B}_1^\dagger &= \bar{\boldsymbol{\Lambda}}\,, \qquad & \boldsymbol{B}_1\boldsymbol{S}_{Y,1}\boldsymbol{B}_1^\dagger &= \boldsymbol{I}\,; \\ \boldsymbol{B}_2\boldsymbol{S}_{X,2}\boldsymbol{B}_2^\dagger &= \bar{\boldsymbol{\Lambda}}\,, \qquad & \boldsymbol{B}_2\boldsymbol{S}_{Y,2}\boldsymbol{B}_2^\dagger &= \boldsymbol{I}\,; \end{aligned}$$

Hence,

$$\begin{aligned} \boldsymbol{S}_{X,1} &= \boldsymbol{B}_1^{-1}\boldsymbol{B}_2\boldsymbol{S}_{X,2}\boldsymbol{B}_2^\dagger(\boldsymbol{B}_1^{-1})^\dagger = \boldsymbol{B}\boldsymbol{S}_{X,2}\boldsymbol{B}^\dagger\,, \\ \boldsymbol{S}_{Y,1} &= \boldsymbol{B}_1^{-1}\boldsymbol{B}_2\boldsymbol{S}_{Y,2}\boldsymbol{B}_2^\dagger(\boldsymbol{B}_1^{-1})^\dagger = \boldsymbol{B}\boldsymbol{S}_{Y,2}\boldsymbol{B}^\dagger\,, \end{aligned}$$

with $\boldsymbol{B} = \boldsymbol{B}_1^{-1}\boldsymbol{B}_2$, and the proof is thus completed.

Interestingly the principle of invariance realizes a significant data reduction: the maximal invariant statistic is a real N-dimensional vector whereas the original sufficient statistic is composed of the two $N \times N$ Grammian matrices $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$. For instance, with reference to three polarimetric channels, the original sufficient statistic is composed of 18 real entries whereas in the compressed domain (i.e. after reduction by invariance) the maximal invariant is just 3-dimensional.

1.2.2 Induced Maximal Invariant Design

The data transformation induces a parameter transformation which leaves unaltered the two composite hypotheses. In other words, by the principle of invariance, the parameter space is also partitioned into orbits and, usually, involves a reduced set of parameters. The relevant parameters are embodied into any *induced maximal invariant*, namely any function of the parameters that is constant over each orbit of the parameter space (invariance) but assumes different values over different orbits (maximality).

For the case at hand, an induced maximal invariant is composed of the eigenvalues $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_N]^T$ of the matrix:

$$\boldsymbol{\Sigma}_X \boldsymbol{\Sigma}_Y^{-1} . \tag{1.9}$$

The proof of this claim can be done following the same steps as in the proof of Proposition 1 and is omitted for the sake of compactness.

The previous claim highlights that the principle of invariance yields a significant reduction of the number of the parameters: in fact, the induced maximal invariant is an N-dimensional vector while the original parameter space was composed of the two covariance matrices $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$.

We explicitly observe that in the reduced parameter space the partition corresponding to the two composite hypotheses of the test (1.1) is $\Xi_0 = \{\mathbf{1}_N\}$, relative to $\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y$, and $\Xi_1 = \overline{\{\mathbf{1}_N\}}$, relative to $\boldsymbol{\Sigma}_X \neq \boldsymbol{\Sigma}_Y$, where $\overline{\{\mathbf{1}_N\}}$ is the set of the N-dimensional column vectors with at least one entry different from 1. The structure of Ξ_0, which now corresponds to a simple H_0 hypothesis, clearly shows that all invariant receivers that process a maximal invariant statistic through a transformation independent of $\delta_1, \ldots, \delta_N$, achieve the CFAR property.

1.3 Design of the optimum and sub-optimum Invariant Detectors

This section is devoted to the design of the optimum and sub-optimum invariant detectors. Sub-section A deals with the design of the MPI detector for the problem of interest and the existence of a UMPI test. Since no UMPI detector exists, Sub-section B introduces some sub-optimum invariant architectures.

1.3.1 Design of the Most Powerful Invariant Detector

Based on the Neyman-Pearson criterion, the MPI test can be obtained as the LRT based on any maximal invariant [16, 29], which, as already pointed out, provides sufficient information for constructing any invariant decision rule. It is thus necessary, for further developments, to introduce the pdf of the devised maximal invariant under both the H_0 and the H_1 hypothesis [26]:

$$f_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}|H_1,\boldsymbol{\Delta}) = c_{N,K}\det(\boldsymbol{\Delta})^{-K}\,{}_1\tilde{F}_0(2K;-\boldsymbol{\Delta}^{-1},\boldsymbol{\Lambda})\det(\boldsymbol{\Lambda})^{K-N}\prod_{j=1}^{N}\prod_{i<j}(\lambda_i-\lambda_j)^2\,, \tag{1.10}$$

$$f_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}|H_0) = f_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}|H_1,\boldsymbol{I}) = c_{N,K}\det(\boldsymbol{I}+\boldsymbol{\Lambda})^{-2K}\det(\boldsymbol{\Lambda})^{K-N}\prod_{j=1}^{N}\prod_{i<j}(\lambda_i-\lambda_j)^2\,, \tag{1.11}$$

where[3] $\boldsymbol{\Lambda} = \mathbf{Diag}\left([\lambda_1,\lambda_2,\ldots,\lambda_N]^T\right)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > 0$, $\boldsymbol{\Delta} = \mathbf{Diag}\,(\boldsymbol{\delta})$, ${}_1\tilde{F}_0(\cdot;\cdot,\cdot)$ denotes the hypergeometric function of matrix argument [26, eq. (88)], and $c_{N,K}$ is a normalization constant.

Hence, the LRT can be written as

$$\frac{f_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}|H_1,\boldsymbol{\Delta})}{f_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}|H_0)} = \frac{{}_1\tilde{F}_0(2K;-\boldsymbol{\Delta}^{-1},\boldsymbol{\Lambda})}{\det(\boldsymbol{I}+\boldsymbol{\Lambda})^{-2K}} \underset{H_0}{\overset{H_1}{\gtrless}} T\,, \tag{1.12}$$

[3] With a slight notation abuse we continue to indicate with λ_i, $i = 1,\ldots,N$, the ordered (in descending order) eigenvalues.

where T denotes the detection threshold, set to ensure a given P_{fa} level.

In order to proceed further, we focus on the case $N = 2$ and exploit expression [30, eq. (51)] to recast ${}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda})$ as a rational function

$$
\begin{aligned}
&{}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda}) = \\
&\qquad \frac{c_2 \det \begin{bmatrix} \left(1+\frac{\lambda_1}{\delta_1}\right)^{1-2K} & \left(1+\frac{\lambda_1}{\delta_2}\right)^{1-2K} \\ \left(1+\frac{\lambda_2}{\delta_1}\right)^{1-2K} & \left(1+\frac{\lambda_2}{\delta_2}\right)^{1-2K} \end{bmatrix}}{(\lambda_1-\lambda_2)\left(-\frac{1}{\delta_1}+\frac{1}{\delta_2}\right)}, \quad \delta_i \neq \delta_j \,,
\end{aligned}
\tag{1.13}
$$

with c_2 a constant with respect to λ_i and δ_i.

Moreover for $\delta_1 = \delta_2 = \delta$, the splitting formula [26, eq. (92)], yields

$$
{}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda}) = \det\left(\boldsymbol{I} + \frac{\boldsymbol{\Lambda}}{\delta}\right)^{-2K}, \quad \delta_i = \delta_j = \delta \,. \tag{1.14}
$$

Otherwise stated,

$$
\begin{aligned}
&{}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda}) = \\
&\begin{cases}
\dfrac{c_2\left[\left(1+\frac{\lambda_1}{\delta_1}\right)^{1-2K}\left(1+\frac{\lambda_2}{\delta_2}\right)^{1-2K} - \left(1+\frac{\lambda_1}{\delta_2}\right)^{1-2K}\left(1+\frac{\lambda_2}{\delta_1}\right)^{1-2K}\right]}{(\lambda_1-\lambda_2)\left(-\frac{1}{\delta_1}+\frac{1}{\delta_2}\right)} & \delta_i \neq \delta_j \\
\left(1+\frac{\lambda_1}{\delta}\right)^{-2K}\left(1+\frac{\lambda_2}{\delta}\right)^{-2K} & \delta_1 = \delta_2 = \delta
\end{cases}
\end{aligned}
\tag{1.15}
$$

As a consequence, after standard statistical equivalences, the MPI test can be recast as

$$\left[\frac{(1+\lambda_1)}{(\delta_1+\lambda_1)}\frac{(1+\lambda_2)}{(\delta_2+\lambda_2)}\right]^{2K}\left[\frac{(\delta_1+\lambda_1)(\delta_2+\lambda_2)}{\lambda_1-\lambda_2}\right]+$$
$$-\left[\frac{(1+\lambda_1)}{(\delta_2+\lambda_1)}\frac{(1+\lambda_2)}{(\delta_1+\lambda_2)}\right]^{2K}\left[\frac{(\delta_2+\lambda_1)(\delta_1+\lambda_2)}{\lambda_1-\lambda_2}\right]\times \tag{1.16}$$
$$\times\,\mathrm{sign}(\delta_1-\delta_2)\underset{H_0}{\overset{H_1}{\gtrless}} T_A\,,\qquad \delta_1\neq\delta_2$$

$$\frac{(1+\lambda_1)(1+\lambda_2)}{(\delta+\lambda_1)(\delta+\lambda_2)}\underset{H_0}{\overset{H_1}{\gtrless}} T_B\,,\qquad \delta_1=\delta_2=\delta \tag{1.17}$$

with T_A and T_B suitable modifications of the original detection threshold in (1.12).

The derived expression of the MPI test clearly highlights that it cannot, in general, be implemented unless the induced maximal invariant is known (clairvoyant detector): unfortunately, such hypothesis is not realistic, especially for the SAR change detection problem. In other words, the MPI detector, which provides the best performance for a given $\boldsymbol{\Delta}^\star$ is not, in general, optimum for different values of $\boldsymbol{\Delta}$, namely no UMPI detector exists. Nevertheless the MPI detector is still noteworthy since its performance upper bounds those of any other invariant receiver, operating under the same signal and disturbance models, and hence can be used to assess the loss of any implementable, although sub-optimal, invariant test. Before concluding, we highlight that the case $N = 3$ can be dealt with a similar technique (the derivations are given in Appendix A).

1.3.2 Maximal Invariant-based Detectors

The lack of an UMPI test suggests to investigate invariant decision rules based upon different strategies. However, there is no criterion for choosing *a priori* a receiver instead of another. An intuitive rule to select invariant tests for our problem could be based on the following asymptotic observation. For large values of K, the eigenvalues of $\boldsymbol{S}_X^{-1}\boldsymbol{S}_Y$ tend to δ_i, $i = 1, \ldots, N$; hence decision rules

$$h(\lambda_1, \ldots, \lambda_N) \underset{H_0}{\overset{H_1}{\gtrless}} T\,, \tag{1.18}$$

1. are very effective to discriminate deviations $\delta_i \gg 1$, when $h(\cdot, \ldots, \cdot)$ is an increasing function of the arguments. However they perform poor when δ_i are smaller than 1.

2. are very effective to discriminate deviations $\delta_i \ll 1$, when $h(\cdot, \ldots, \cdot)$ is a decreasing function of the arguments. Nevertheless they perform poor when δ_i is greater than 1.

3. in principle could achieve good detection levels for both $\delta_i \gg 1$ and $\delta_i \ll 1$, when $h(\cdot, \ldots, \cdot)$ complies with $h\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_N}\right) = h\left(\lambda_1, \ldots, \lambda_N\right)$.

On the other hand one cannot analyze all possible reasonable detectors; so in the following, we focus on six decision rules, which, based on extensive numerical analysis, are seen to achieve satisfactory detection performance.

GLRT

This approach is equivalent to replacing the unknown parameters in the likelihood ratio with their maximum likelihood estimates, under each hypothesis [31]. Interestingly, under very mild technical assumptions, the GLRT is invariant [32]. For the present problem it has been proposed in [8, 9][4] and is given by

$$\frac{\max\limits_{\boldsymbol{\Sigma}_X}\max\limits_{\boldsymbol{\Sigma}_Y} f_{\boldsymbol{S}_X,\boldsymbol{S}_Y}(\boldsymbol{S}_X, \boldsymbol{S}_Y | H_1, \boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y)}{\max\limits_{\boldsymbol{\Sigma}_X} f_{\boldsymbol{S}_X,\boldsymbol{S}_Y}(\boldsymbol{S}_X, \boldsymbol{S}_Y | H_0, \boldsymbol{\Sigma}_X)} \underset{H_0}{\overset{H_1}{\gtrless}} T_1\,, \tag{1.19}$$

with T_1 the detection threshold. After optimizations and monotonic transformations, (1.19) can be shown statistically equivalent to

$$\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i} \underset{H_0}{\overset{H_1}{\gtrless}} T_1\,, \tag{1.20}$$

[4] With reference to real observations it is derived in [15, Ch. 8].

where the same symbol T_1 has been used to denote the modified detection threshold. Interestingly, the GLRT complies with condition 3 that was given at the beginning of the section, namely we expect the GLRT capable to achieve good detection levels both when $\delta_i \gg 1$ and $\delta_i \ll 1$.

Arithmetic and Armonic Mean Based Detectors

These decision rules are respectively given by

$$\sum_{i=1}^{N} \lambda_i \underset{H_0}{\overset{H_1}{\gtrless}} T_2\,, \tag{1.21}$$

$$\sum_{i=1}^{N} \frac{1}{\lambda_i} \underset{H_0}{\overset{H_1}{\gtrless}} T_3\,, \tag{1.22}$$

where T_2 and T_3 are the detection thresholds. The former complies with condition 1 whereas the latter with condition 2. As a consequence (1.21)[5] is suitable to detect deviations $\delta_i \gg 1$ while (1.22) $\delta_i \ll 1$.

From (1.21) and (1.22) it is also possible to construct another decision rule satisfying condition 3 merely summing the decision statistics, i.e.

$$\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i \right) \underset{H_0}{\overset{H_1}{\gtrless}} T_4\,, \tag{1.23}$$

with T_4 the detection threshold. Before concluding this paragraph we highlight that (1.21) can be also obtained substituting in the optimum LRT detector (1.5) the sample covariance matrices $\frac{1}{K}\boldsymbol{S}_X$ and $\frac{1}{K}\boldsymbol{S}_Y$ in place of $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ respectively. Otherwise stated, the arithmetic mean based detector can be interpreted as an adaptive implementation of the LRT.

Maximum and Minimum Eigenvalue Based Detectors

These tests are respectively based on the following comparisons

$$\left(\lambda_1 + \frac{1}{\lambda_N} \right) \underset{H_0}{\overset{H_1}{\gtrless}} T_5\,, \tag{1.24}$$

[5] This detector is also known as The Hotelling-Lawley trace and has been discussed for change detection purposes in [33].

$$\max\left(\lambda_1\,,\,\frac{1}{\lambda_N}\right)\underset{H_0}{\overset{H_1}{\gtrless}}T_6\,, \tag{1.25}$$

with T_5 and T_6 the decision thresholds. An intuitive explanation to the decision rules is based on the following arguments: the former term, λ_1, dominates for large deviations $\delta_i \gg 1$, whereas, the latter term, $\frac{1}{\lambda_N}$, if $\delta_i \ll 1$. Hence, (1.24) and (1.25) are supposed to perform well both for $\delta_i \gg 1$ and $\delta_i \ll 1$.

1.4 Performance Analysis on Simulated Data

This section presents the performance analysis via computer simulated data of the detectors introduced in Section IV. It is organized in two sub-sections. The first analyzes P_d, for a given P_{fa}, versus the induced maximal invariant which represents the set of relevant parameters ruling the P_d's behavior (otherwise stated, different pairs of covariance matrices $(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y)$ sharing the same induced maximal invariant $\boldsymbol{\delta}$ lead to the same detection performance). The second shows standard ROCs for a given value of $\boldsymbol{\delta}$. Finally, the impact of the number of polarimetric channels used to perform change detection is studied.

1.4.1 Detection Probability versus the Induced Maximal Invariant

The analysis is conducted in terms of P_d for a given P_{fa} level, assuming zero-mean complex circular multivariate Gaussian observations with equal (distinct) covariance matrices under H_0 (H_1). For comparison purposes, the MPI test is used as a benchmark to the performance of any feasible invariant receiver. Finally a square inspection window is considered, namely $W_1 = W_2 = W$.

The case of two polarizations is chosen because for $N = 2$ the induced maximal invariant, which rules the performance of any invariant detector, is bi-dimensional. This means that contour plots of P_d versus the components of the induced maximal invariant δ_1 and δ_2 completely characterize the detection performance of receivers (1.20) - (1.25). Specifically, the curves can be interpreted as follows. Given a pair of true covariance matrices $(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y)$, compute the bi-dimensional induced maximal

invariant $\boldsymbol{\delta}$, and from the countours read the value of P_d corresponding to a given P_{fa} level.

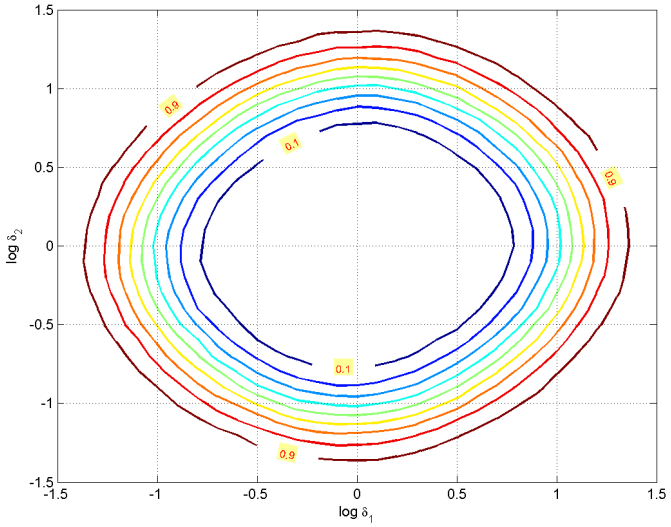
In the present analysis, Monte Carlo simulation is used to set the detection thresholds assuming $100/P_{fa}$ independent runs and $P_{fa} = 10^{-4}$. Moreover, 5000 independent trials are used to estimate P_d.

Figures 1.2, 1.3, and 1.4, show the contours corresponding to $W = 3$. They highlights that receivers complying with condition 3 in sub-section 1.3.2 are capable of detecting deviations δ_i either $\gg 1$ or $\delta_i \ll 1$. Moreover, (1.23), (1.24), and (1.25) achieve a comparable detection performance level, namely $P_d \geq 0.9$ for $\delta_i \notin [10^{-1.3}\,,\, 10^{1.3}]$, $i = 1, 2$, while the same level of performance is achieved by (1.20) for $\delta_i \notin [10^{-1.36}\,,\, 10^{1.36}]$, $i = 1, 2$.

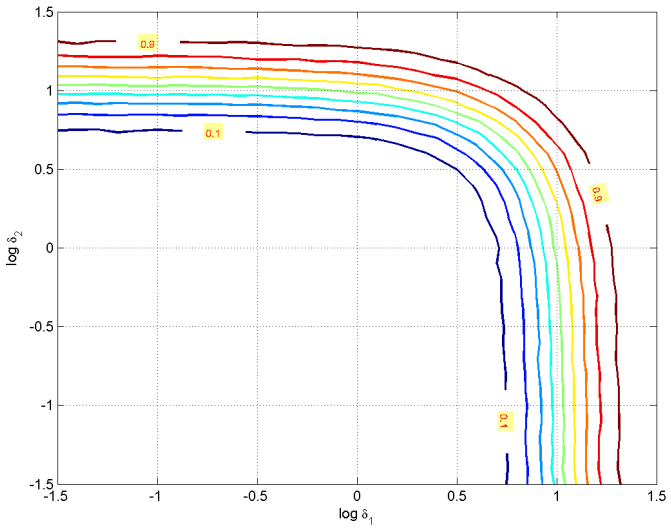
A different behavior is shown by tests (1.21) and (1.22); the former operates with a satisfactory detection performance when $\delta_i \notin [0\,,\, 10^{1.3}]$, $i = 1, 2$, whereas the latter when $(\delta_1, \delta_2) \in]0\,,\, 10^{-1.3}]^2$. This is a useful feature as, it will be shown on real data, tests (1.21) and (1.22) can be used for post-detection classification, namely to discriminate between $\delta_i \gg 1$ and $\delta_i \ll 1$.

In Figures 1.5 and 1.6, the performance of the considered invariant tests is compared with the benchmark MPI receiver. For this test $W = 3$ and four different cuts of the detection probability surface (P_d versus δ_1 and δ_2) are considered: $\delta_2 = \delta_1$, $\delta_2 = 0.75\delta_1$, $\delta_2 = 0.50\delta_1$, and $\delta_2 = 0.25\delta_1$. The curves highlight that tests (1.21) and (1.22) respectively optimized to the specific situation $\delta \gg 1$ (1.21) and $\delta \ll 1$ (1.22) slightly outperform the other detection rules; additionally, their performance is near to the optimum curve in the region $\delta_1 > 10$ (1.21) and $\delta_1 < 0.1$ (1.22). However, their P_d is close to zero when $\delta_1 < 10$ (1.21) and $\delta_1 > 0.1$ (1.22). All the remaining receivers follow the same performance behavior as the benchmark curve.

The case $W = 5$ is not reported for conciseness, however it highlights that increasing W leads to better detection performances. Moreover, the performance gap between the considered invariant detectors and the MPI test is smaller than that observed for $W = 3$. This was expected and can be explained observing that a greater number of homogeneous data vectors have been used to construct the Grammians $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ whose scaled versions ($\frac{1}{K}$ scale factor) are unbiased and consistent estimates of the covariance matrices $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$.

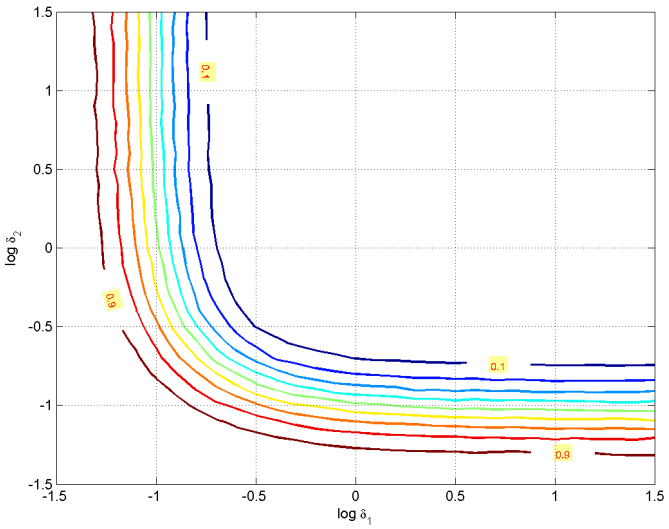


(a) Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$.

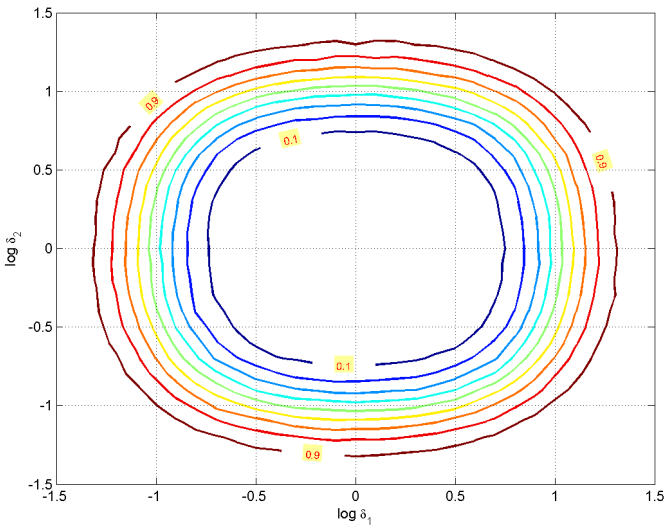


(b) Detector (1.21): $\sum_{i=1}^{N} \lambda_i$.

Figure 1.2: P_d contours versus $\log \delta_1$ and $\log \delta_2$ for $N = 2$ and $W = 3$.

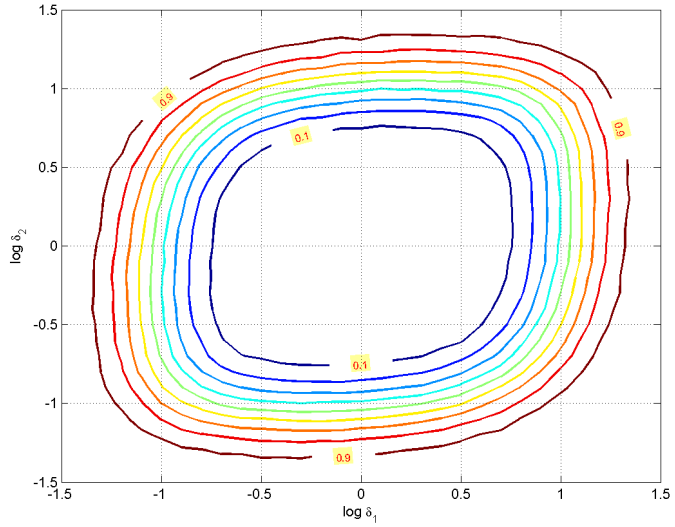


(a) Detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$.

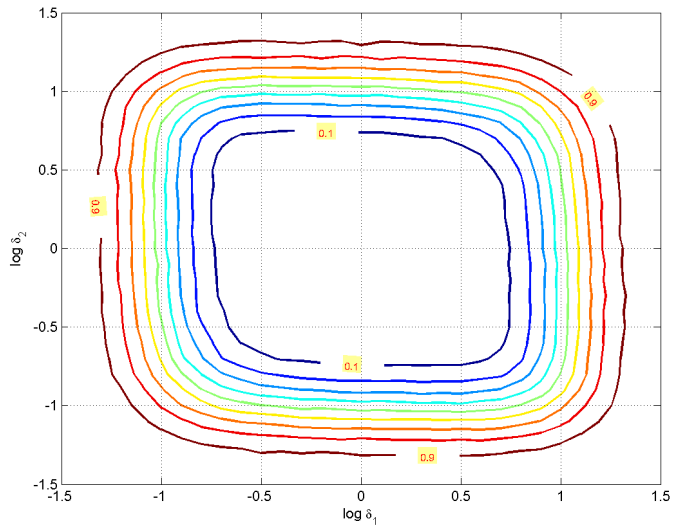


(b) Detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i \right)$.

Figure 1.3: P_d contours versus $\log \delta_1$ and $\log \delta_2$ for $N = 2$ and $W = 3$.

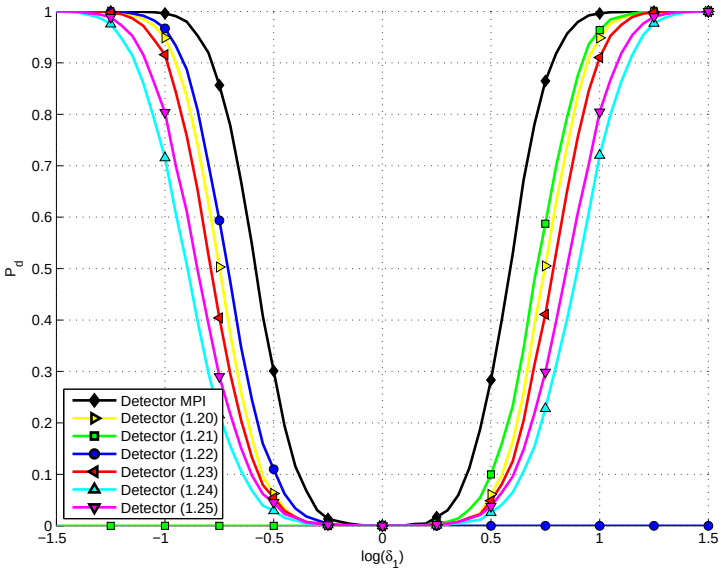


(a) Detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$.

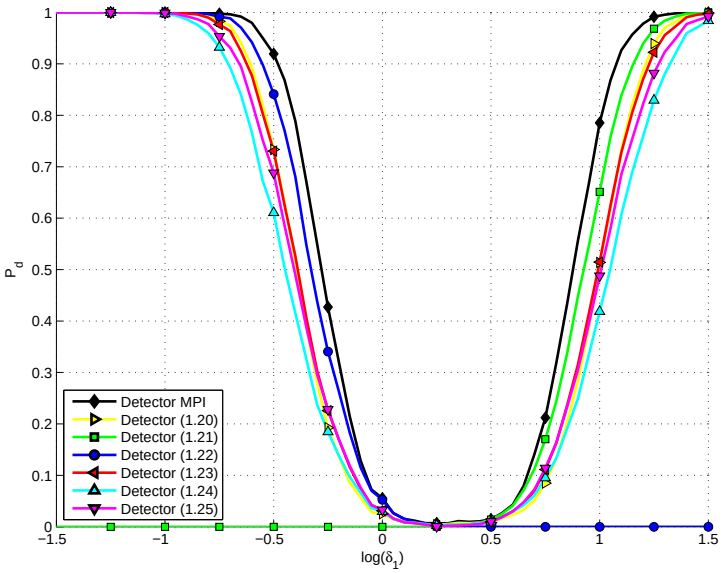


(b) Detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$.

Figure 1.4: P_d contours versus $\log \delta_1$ and $\log \delta_2$ for $N = 2$ and $W = 3$.

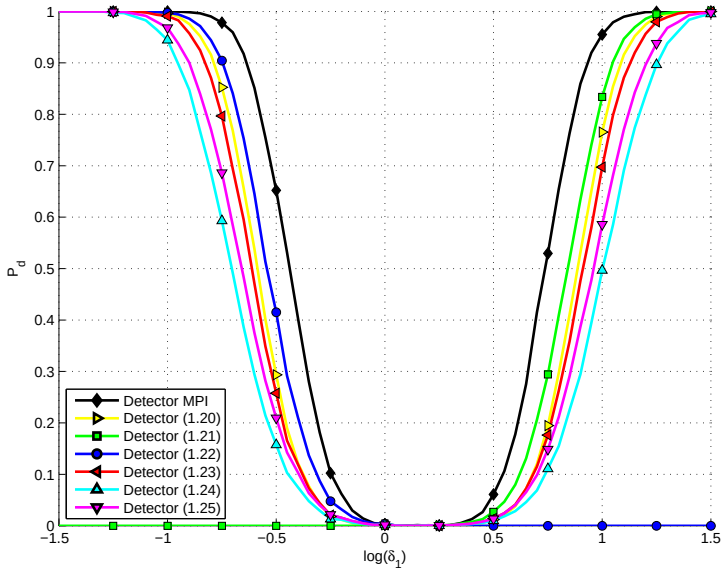


(a) $\delta_1 = \delta_2$.

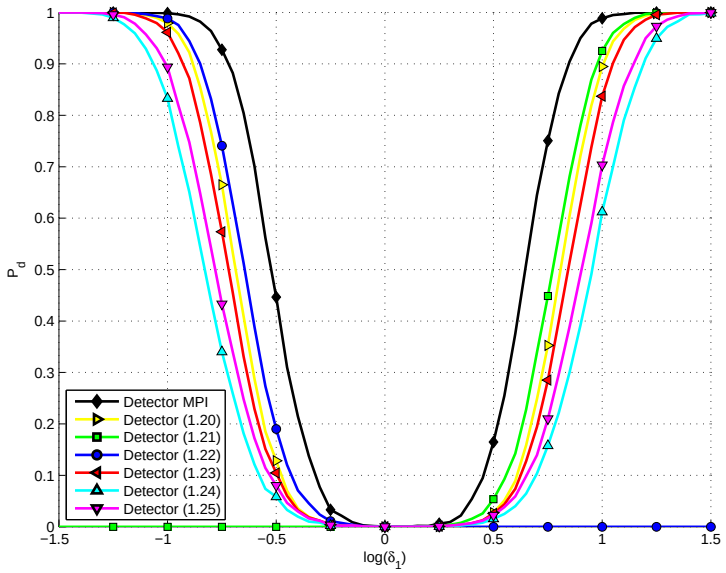


(b) $\delta_1 = 0.25\delta_2$.

Figure 1.5: P_d versus $\log \delta_1$ for $N = 2$ and $W = 3$. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$.



(a) $\delta_1 = 0.5\delta_2$.



(b) $\delta_1 = 0.75\delta_2$.

Figure 1.6: P_d versus $\log \delta_1$ for $N = 2$ and $W = 3$. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1, \frac{1}{\lambda_N}\right)$.

1.4.2 Standard ROC Analysis

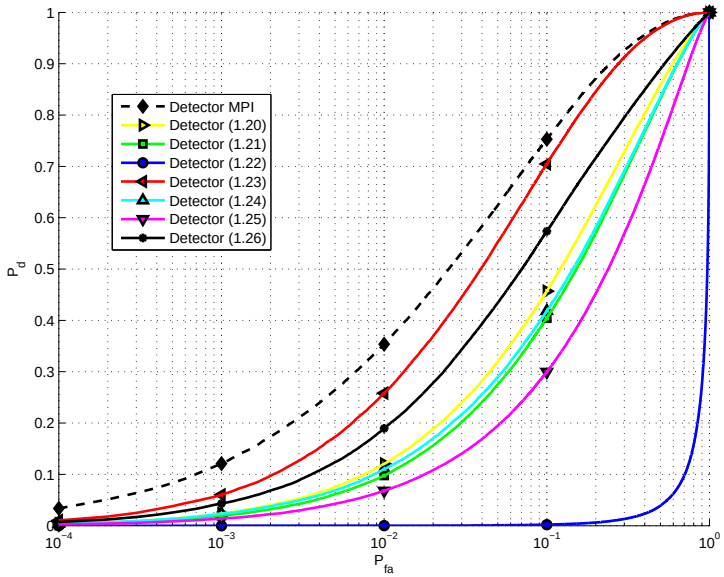
The analysis is conducted in terms of ROC curves (namely P_d versus P_{fa}) for a fixed value of the induced maximal invariant $\boldsymbol{\delta} = [1/2, 1/2, 1/2]^T$, corresponding to the pair of covariance matrices exploited in [9]. Of course, any other covariance pair with the same value of $\boldsymbol{\delta}$ leads to the same ROCs, while a different value of $\boldsymbol{\delta}$ leads to different ROCs. For comparison purposes, the benchmark ROC of the clairvoyant MPI detector and that of the detector proposed in [9] are also plotted. The detector proposed in [9] is an adaptive implementation of the LRT, whose expression in terms of the maximal invariant statistic is

$$\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i} \right) \underset{H_0}{\overset{H_1}{\gtrless}} T_H \, . \tag{1.26}$$

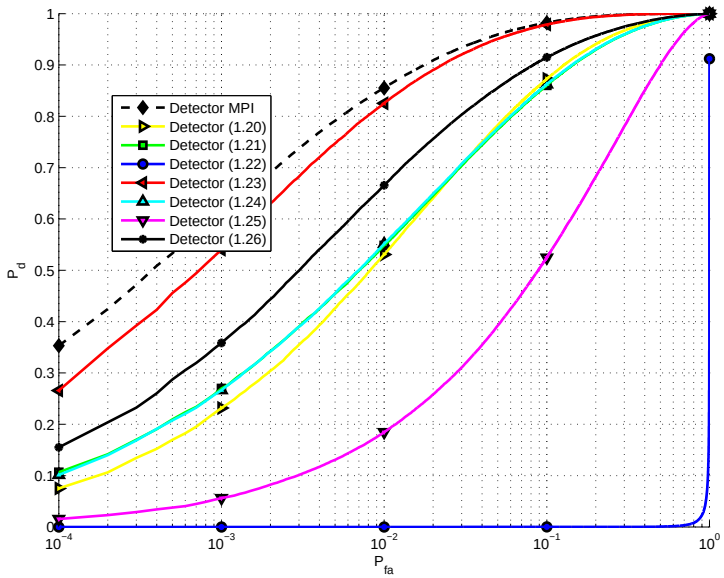
In order to set the detection threshold, Monte Carlo simulations are used assuming $100/P_{fa}$ independent runs. In the simulations 5000 independent trials were used to estimate the P_d. Figures 1.7 and 1.8 show, respectively, the ROCs of the considered receivers for two[6] and three polarimetric channels considering the cases $W = 3$ and $W = 5$.

In all the analyzed cases, the ROC highlights that the Armonic Mean based Detector (1.22) outperforms the counterparts and the loss with respect to the benchmark curve is acceptable for almost all the parameters values chosen in the simulation. The GLRT (1.20) and the detector (1.23) achieves almost the same performance levels and come third and are ranked just after detector (1.26). Considering the Arithmetic Mean based Detector, its P_d for the given value of $\boldsymbol{\delta}$ is completely unsatisfactory which is not surprising and is actually anticipated by the theory, since the decision statistic belongs to the class 1) of subsection 1.3.2. Summarizing, the simulation curves evidently show that the GLRT (1.20) is often outperformed by the counterparts and, as predicted by the theory, there is not a detector which uniformly outperforms the others.

[6] For the covariance pairs in [9] the HH, VV, and HV polarimetric channels are perfectly equivalent because the induced maximal invariant exhibits equal components.

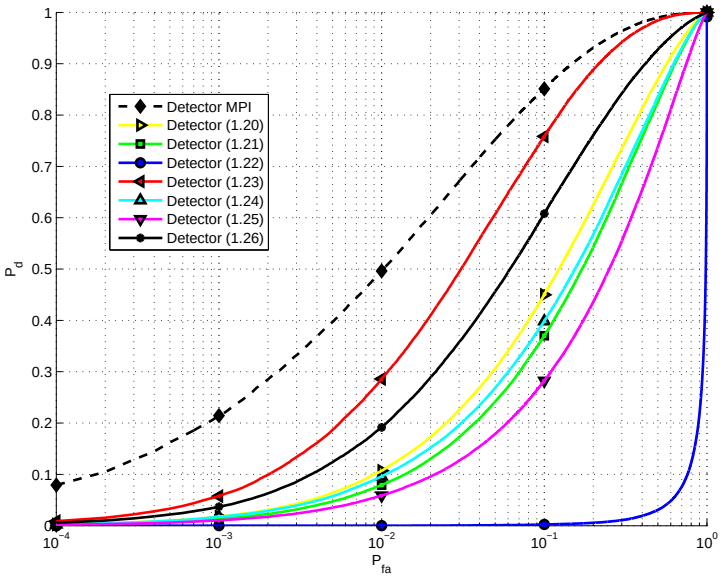


(a) $W = 3$.

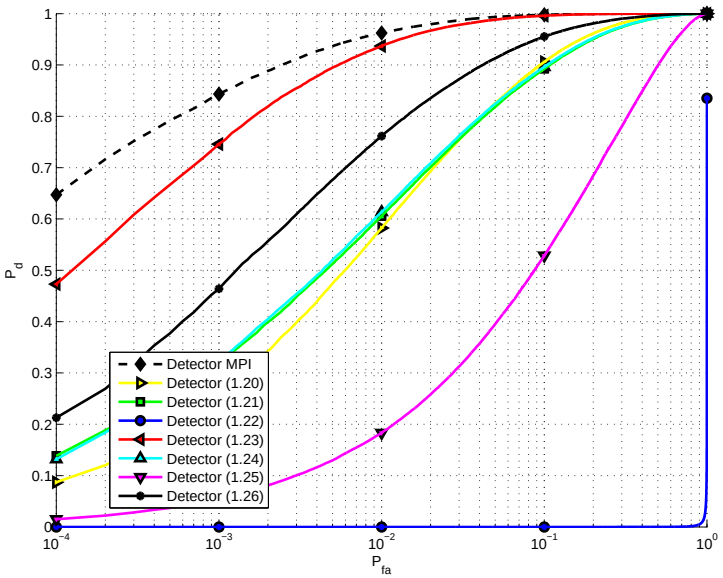


(b) $W = 5$.

Figure 1.7: P_d versus P_{fa} considering $N = 2$. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$, detector (1.26): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$.



(a) $W = 3$.



(b) $W = 5$.

Figure 1.8: P_d versus P_{fa} considering $N = 3$. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$, detector (1.26): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$.

The effects on the performance of the number of processed polarimetric channels are studied in Figures 1.9 and 1.10. For conciseness the analysis is limited to the ROC curves of detectors (1.20), (1.22), (1.25) and (1.26), which are reported for $W = 5$, $N = 3$, $N = 2$, and $N = 1$. Based on footnote 4, the curves for $N = 1$ can be either interpreted as the ROC of the HH, VV, or the HV channel. It is also worth mentioning that for $N = 1$ no UMPI detector exists. In fact, for $N = 1$, the MPI receiver

$$\frac{\boldsymbol{S}_X}{\boldsymbol{S}_Y}\text{sign}(\delta - 1) \underset{H_0}{\overset{H_1}{\gtrless}} T_{A,1}\,,$$

requires the knowledge of the induced maximal invariant $\delta = \boldsymbol{\Sigma}_X/\boldsymbol{\Sigma}_Y$.

The analysis of the curves in Figures 1.9 and 1.10 shows that the detection performance for a given P_{fa} level improves as N increases. For example, with reference to detector (1.20) and a P_{fa} of 10^{-3}, a P_d of 0.18 is achieved for $N = 1$, while for $N = 2$ and $N = 3$ the P_d results to be 0.27 and 0.32 respectively. This is in accordance with the intuitive observation that the higher the polarimetric information about the scene, the better the performance.

So far, all the numerical analysis performed in this sub-section was based on the induced maximal invariant value $\boldsymbol{\delta} = [1/2, 1/2, 1/2]^T$. Thus, in order to provide an average behavior of the ROC curves for the considered detectors, the following simulation is conducted:

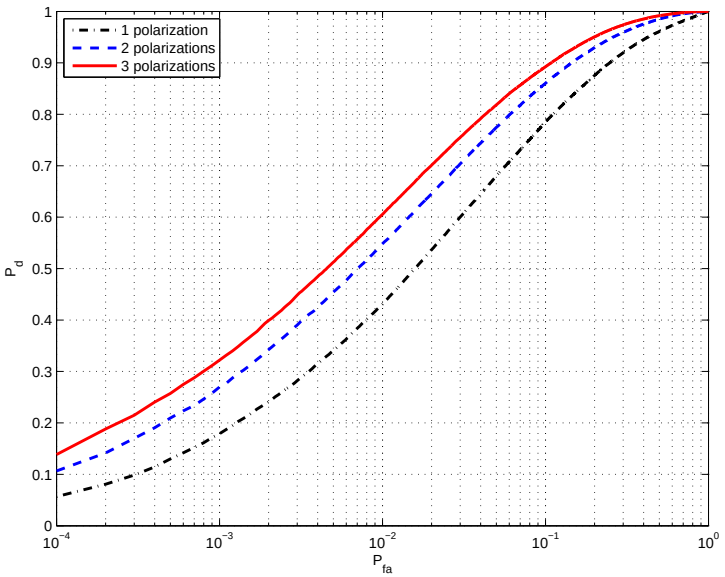
1. 500 random values[7] of the induced maximal invariant are selected (modeling the entries of $\boldsymbol{\delta}$ as independent and identically distributed uniform random variables within the intervals $(0, 1)$, $(1, 2)$ and $(0, 2)$). The corresponding three simulation setups will be referred to in the following as case 1, case 2, and case 3;

2. the ROC curve of a given detector is evaluated in correspondence of each induced maximal invariant realization;

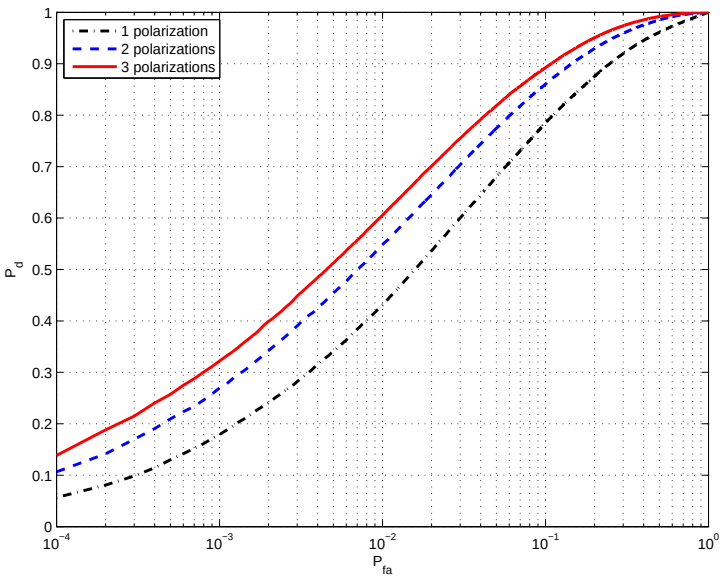
3. for each detector, the previously obtained 500 ROC curves are averaged for a specific simulation setup (i.e. case 1, 2, or 3).

The results are displayed in Figures 1.11, 1.12, and 1.13, where, for conciseness, only the full polarization case is shown.

[7] The seed of the rand MATLAB function is set to the initial state so as to guarantee the reproducibility of the curves.

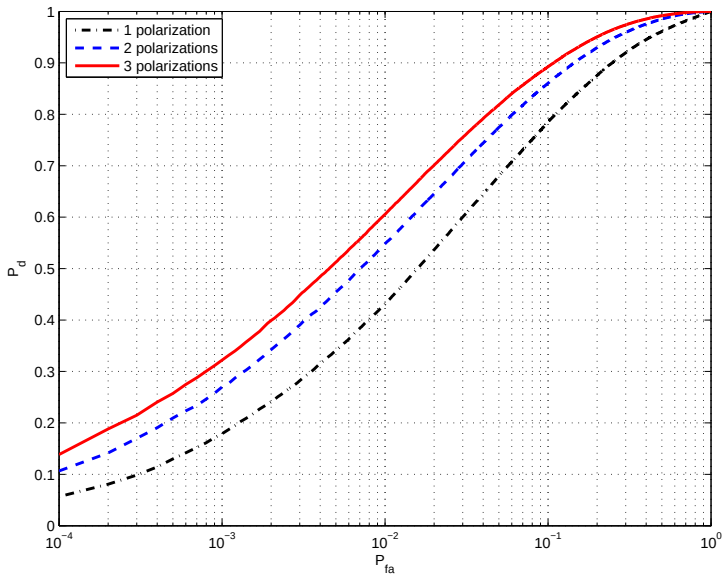


(a) Detector (1.20) $= \prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$.

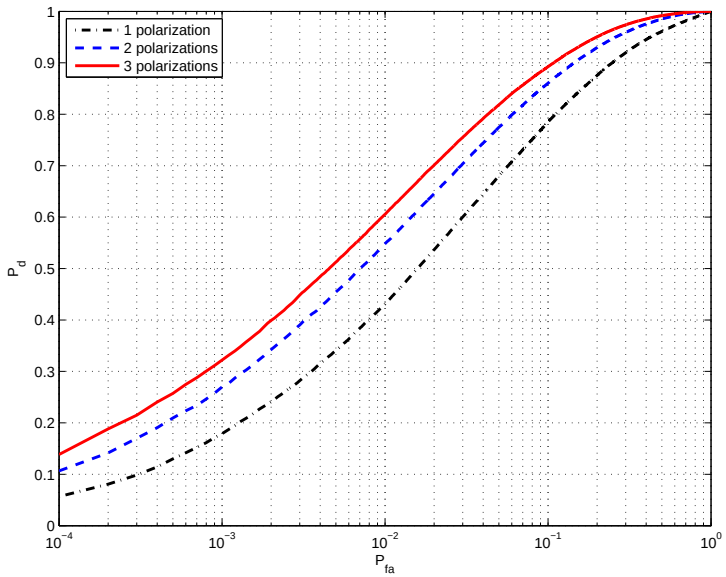


(b) Detector (1.22) $= \sum_{i=1}^{N} \frac{1}{\lambda_i}$.

Figure 1.9: P_d versus P_{fa} for $W = 5$.

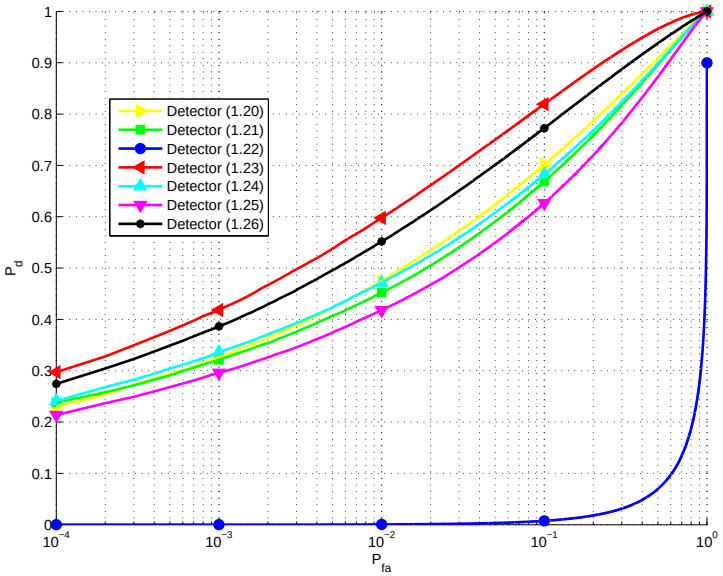


(a) Detector (1.25) $= \max\left(\lambda_1\,,\,\frac{1}{\lambda_N}\right)$.

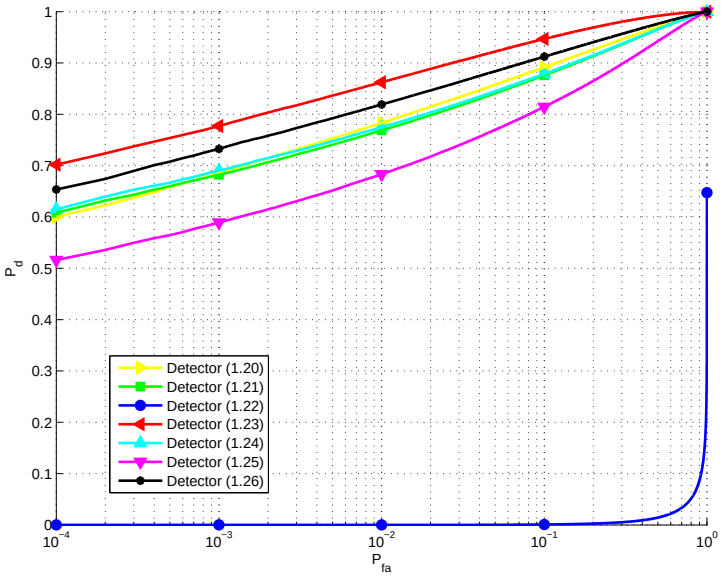


(b) Detector (1.26) $= \sum_{i=1}^{N}\left(\frac{1}{\lambda_i} - \ln\frac{1}{\lambda_i}\right)$.

Figure 1.10: P_d versus P_{fa} for $W = 5$.

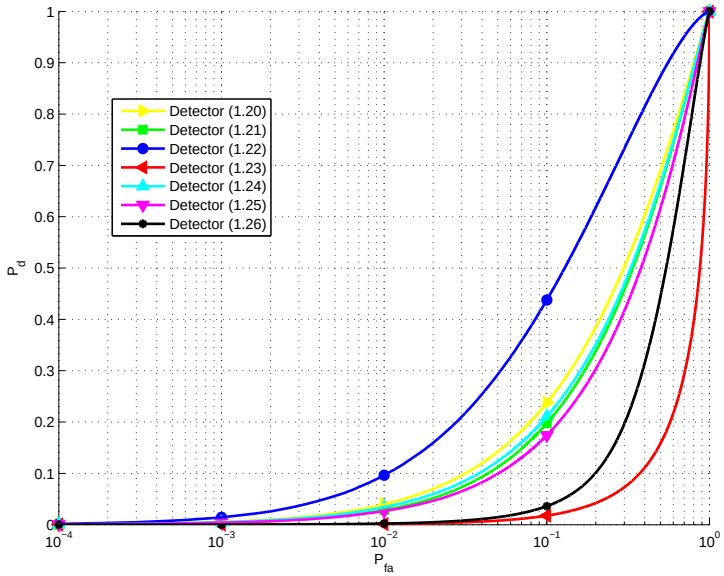


(a) $N = 3$, $W = 3$.

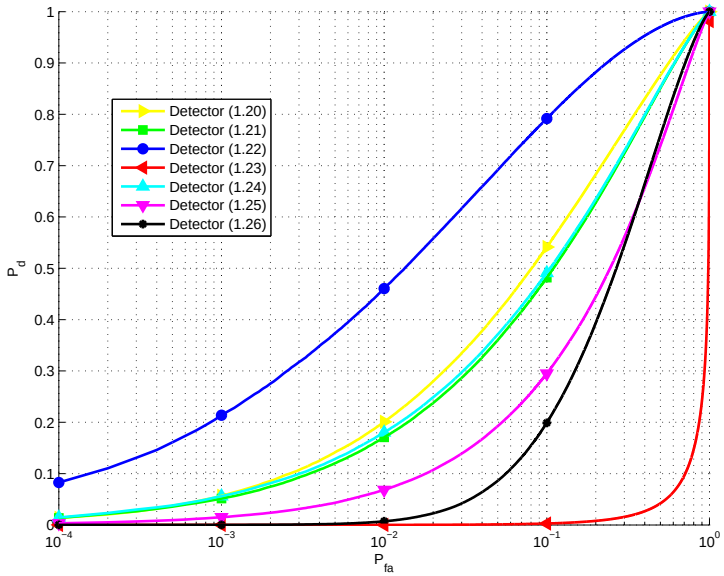


(b) $N = 3$, $W = 5$.

Figure 1.11: Case 1: P_d versus P_{fa}. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$, detector (1.26): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$.

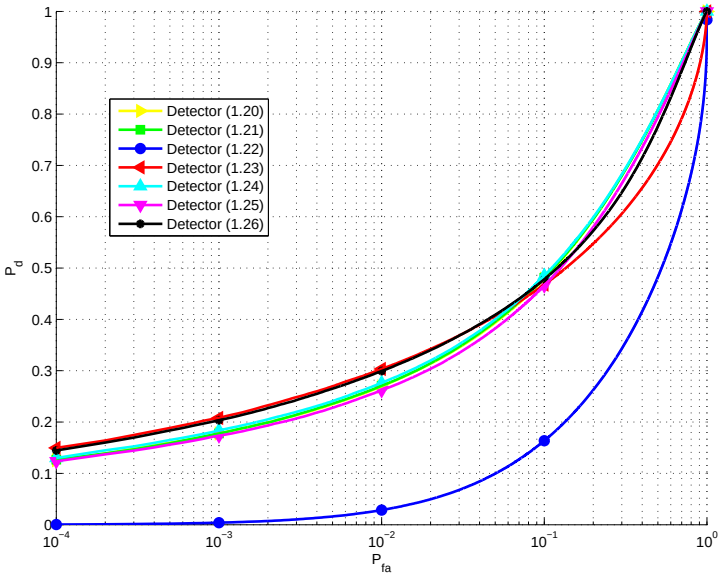


(a) $N = 3$, $W = 3$.

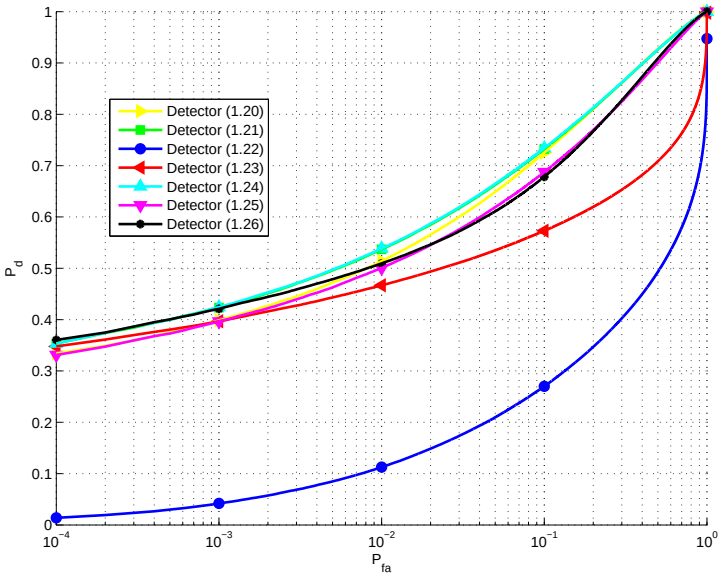


(b) $N = 3$, $W = 5$.

Figure 1.12: Case 2: P_d versus P_{fa}. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$, detector (1.26): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$.



(a) $N = 3$, $W = 3$.



(b) $N = 3$, $W = 5$.

Figure 1.13: Case 3: P_d versus P_{fa}. Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.21): $\sum_{i=1}^{N} \lambda_i$, detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$, detector (1.26): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$.

As expected, when the induced maximal invariant is simulated according to case 1, the ROC curves agree with those shown in Figures 1.7 and 1.8 with detector (1.22) performing better than the counterparts and (1.21) providing the poorest performance.
On the contrary, when case 2 is considered, a different behaviour of the ROC curves is observed. In particular, in the hierarchy detectors (1.21) and (1.22) switched their positions compared to the previous case (this agrees with the theoretical considerations at the beginning of sub-section 1.3.2). Finally, when case 3 is simulated, the ROC curves results in between the two previous situations. This can be explained observing in case 3, the performance can be obtained as the average of that in case 1 and 2.

Summarizing this last analysis confirms that there is not a detector able to able to uniformly outperform the others. This implies that a battery of detectors could be considered with a suitable aggregation logic or some a priori knowledge (if any) about the characteristics of the operating polarimetric environment could be exploited to select the most suitable decision rule. Otherwise stated, based on the operating scenario (urban, sea, foliage) and the targets (car, tank, boats) of tactical interest for change detection, it is possible to exploit a priori models for the polarimetric covariance matrices which allow to gain some information about the expected induced maximal invariant. Based on this last vector, the most suitable detector can be selected with in the class of proposed invariant rules (for instance using the plots of Figures 1.2, 1.3 and 1.4 if $N = 2$).

1.5 Testing on Real Data

In this section the performance analysis of the algorithms proposed in Section 1.3 is presented.

The analysis was performed using real X-band data; the dataset used is the Coherent Change Detection Challenge dataset acquired by the Air Force Research Laboratory (AFRL) [20, 21]. The airborne SAR used to acquire the dataset employed a coherent receiver with 640 MHz bandwidth and dual-polarized mode. The depression angle was of 45 degrees for all the images to scene centre. The data is in the form of focused complex images with a range and cross-range resolution of 0.3 m. The overall dataset provides 10 complex images for each of the three

available polarizations (HH, VV, and HV), acquired the same day. The original images size is of 4501×4501 pixels and have been coherently aligned to a single reference (per polarization) with the help of Digital Elevation Map (DEM) information [20, 21].

For our analysis we focus on two acquisitions from the entire dataset. Unfortunately the ground truths of the data is not available (e.g. the actual changes between two different acquisitions), so the selection of two passes providing the opportunity to generate a sufficiently accurate ground truth is required. Hence two passes satisfying this requirement have been identified: the acquisition named "FP0124" is used as reference pass, while the acquisition "FP0121" is used as a test pass. From the two acquisitions the area with the highest activity (changes) between the images has been selected, this region is represented by a sub-image of 1000×1000 pixels (i.e., $L = M = 1000$) and is composed of several parking lots which are occupied by numerous parked (i.e., stationary) vehicles. Figure 1.14 shows the reference and test sub-images for the HH mode.

For this particular scenario the changes between the reference and test images (denoted by $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively), occurred during the time interval between the two acquisitions can be distinguished in two cases:

- a vehicle is present in $\boldsymbol{X}$ but is not present in $\boldsymbol{Y}$, i.e., the vehicle has departed from its parking space (pixels relative to this kind of event will be referred to in the following as departures);

- a vehicle is not present in $\boldsymbol{X}$ but is present in $\boldsymbol{Y}$, i.e., the vehicle has arrived in an empty parking space (pixels relative to this kind of event will be referred to in the following as arrivals).

Using the cases defined above, a total of 34 changes between $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be visually identified (by flickering the two images). In the analysis the straight line crossing the test image has not been considered, as its nature does not represent an arrival. However, as it is visible in the test image (but not visible in the reference image) we expect it to be detected as a change. The obtained ground truth is shown in Figure 1.15, where the black regions represent the departures and the white ones indicate the arrivals.

(a) Reference image.

(b) Test image.

Figure 1.14: Reference and test images gathered in HH mode.

In particular, denoting by $\mathcal{K}$ the set of pixels that correspond to changes, the ground truth can be represented as a matrix $\boldsymbol{G}$ whose entries are given by

$$\boldsymbol{G}(l,m) = \begin{cases} 1 & \text{if } (l,m) \in \mathcal{K} \quad l = 1,\ldots,L \\ 0 & \text{otherwise} \qquad\quad m = 1,\ldots,M\,. \end{cases} \tag{1.27}$$

In Figure 1.16-a the ground truth mask $\boldsymbol{G}(l,m)$ is shown. Although the acquisitions were performed during the same day and the images were registered, the returns from a scatterer contribute differently to neighbour pixels, for example a slightly different aspect angle can produce a different amount of energy spill-over. These relative differences in the imaged data can lead to false alarms in the change detection results. In order to prevent false alarm caused by pixel contamination by target returns, we consider a guard area around each arrival-departure. This allows the definition of an extended ground truth (see Figure 1.16-b) used in the following to compare the performance of the considered detection algorithms.

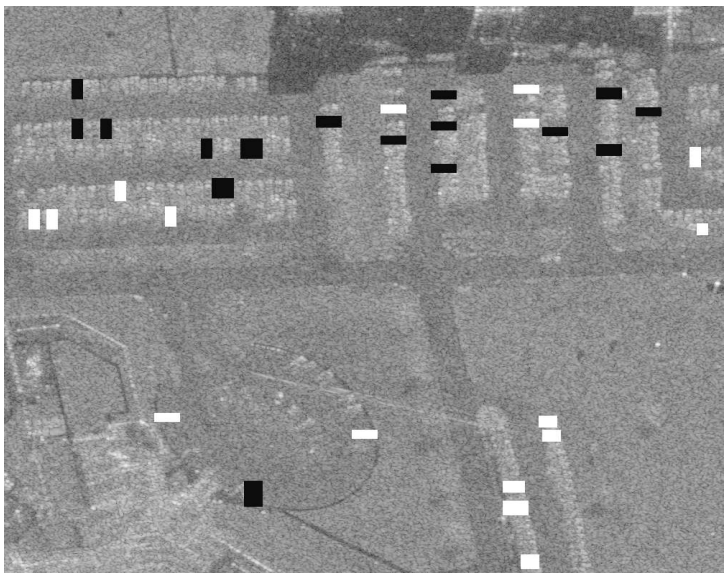
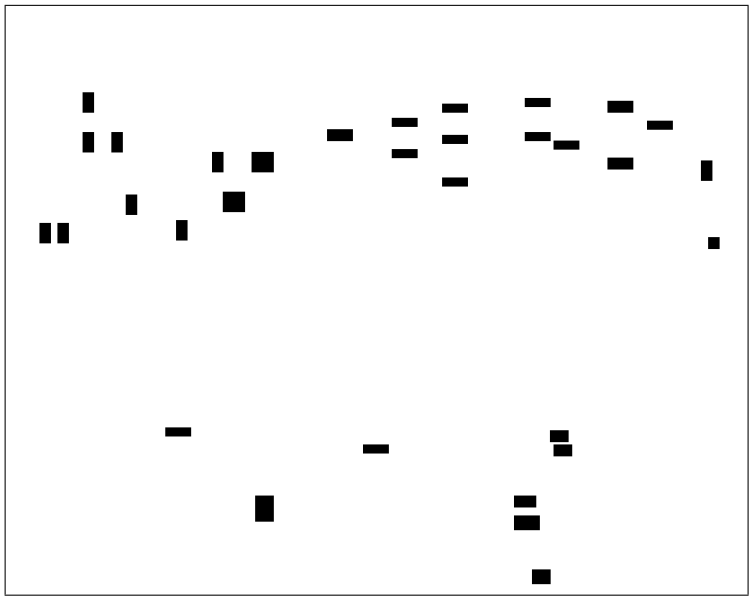
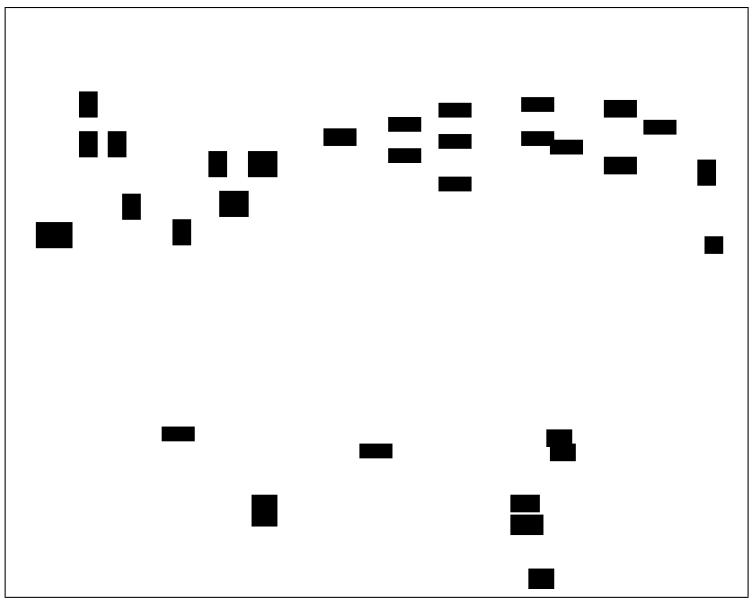


Figure 1.15: Ground truth superimposed to the reference image.



(a) Ground truth.



(b) Ground truth with guard cells.

Figure 1.16: Ground truth without/with the addition of guard cells.

In order to assess the performance of the detectors both the number of detected changes and the change detection maps are presented. For the i-th receiver, the corresponding map of changes $\boldsymbol{C}_i$, is a $L \times M$ matrix whose (l,m)-th entry is the i-th decision statistic considering the $N \times N$ sample Grammian matrix $\boldsymbol{S}_{X_{l,m}}$ ($\boldsymbol{S}_{Y_{l,m}}$) evaluated considering a square neighbourhood[8] with size $W \times W$ of the pixel (l,m) of $\boldsymbol{X}$ ($\boldsymbol{Y}$).

The detection map corresponding to $\boldsymbol{C}_i$, is then defined as

$$\boldsymbol{D}_i(l,m) = \begin{cases} 1 & \text{if } \boldsymbol{C}_i(l,m) > T_i \qquad l = 1,\ldots,L \\ & \\ 0 & \text{otherwise} \qquad\quad m = 1,\ldots,M\,. \end{cases} \tag{1.28}$$

where T_i denotes the detection threshold. In the analysis presented in this section, the thresholds are set to ensure $P_{fa} = 10^{-3}$ in the complement of the extended ground truth area, namely, in the region where no changes occur (there are no true positives). This means that, for each detector, after computing the decision statistics (for each pixel belonging to the complement of the extended ground truth), the threshold has been selected in order that

$$10^{-3} \times \text{total number of available statistics (trials)}\,,$$

are greater than the threshold. This ensures that all the comparisons refer to the same P_{fa} level, namely the number of threshold crossings in the complement of the extended ground truth is exactly the same for all the analyzed detectors.

Finally, are considered the cases of $N = 1$, $N = 2$, and $N = 3$ which correspond, respectively, to process the single polarimetric channel (HH, VV, or HV), to jointly consider a polarization pair (HH-VV, HH-HV, or VV-HV), or to jointly consider all the available channels (HH, VV and HV). In particular, for each detector, starting from $\boldsymbol{D}_i$, and denoting by $\mathcal{K}_G$ the set of the pixels that correspond to changes in the extended ground truth area, performances are given in term of the number of detections belonging to $\mathcal{K}_G$, i.e. considering the cardinality of the set

$$\mathcal{D}_i = \{\boldsymbol{D}_i(l,m) \,:\, \boldsymbol{D}_i(l,m) \in \mathcal{K}_G,\}$$

[8]We notice that, in order to obtain $\boldsymbol{C}_i$ of size $L \times M$ we include a frame of ε pixels width of both reference and test images with $\varepsilon = \frac{W-1}{2}$, in order to be able to compute the statistics on the image borders. By doing so, W must be odd.

where G represents the number of guard cells used to generate the extended ground truth map. In this analysis $G = 5$ is selected, meaning that around each set of pixels representing a change a 5 pixel frame is included.

The number of detections for each receiver corresponding to an actual change present in the extended ground truth are summarized in Table 1.1.

As expected, the common trend is that the performance improves by increasing W. Besides, as to the effects of polarization, the following remarks are in order:

- processing either VV or HV channel leads in general to better performances than processing the HH polarization; the cross-polarized returns provides the highest number of correct detections for the single channel processing (the sole exception is receiver (1.26) for $W = 5$; in this case the VV channel leads to the best choice). The high discrimination capability of the cross-polarized returns can be explained observing that in the presence of a car target there are dihedrals which significantly depolarize the incident radiation. On the contrary the depolarization effect is much limited in the absence of car since the parking area is much more similar to a flat plane. Moreover, the intuitive reason of the better performances achievable with the VV polarization compared with those obtained with the HH one, resides in the dependence of the polarization return on the specific change considered (absence or presence of the car object in the image) and on its geometry (height extension and acquisition geometry).

- with reference to two polarimetric channels, jointly processing the pair VV-HV seems the best choice for almost all the considered decision rules (with exceptions represented by detectors (1.21), for $W = 5$, for which the HH-VV pair is the most suitable choice, and detector (1.26), for $W = 9$, for which polarization diversity does not seem to provide detection improvements).

- a suitable two channel polarization diversity can outperform the single channel processing (the sole exception is detector (1.26) for $W = 9$);

- processing three polarimetric channels does not always improve

the results obtained using two polarizations; the possible improvement depends both on the parameter W as well as on the specific decision rule. In particular for $W = 9$, it appears more convenient the use of a polarization diversity of order two with the pair VV-HV (the sole exception is detector (1.21) whose correct detections increases exploiting the full polarimetric information).

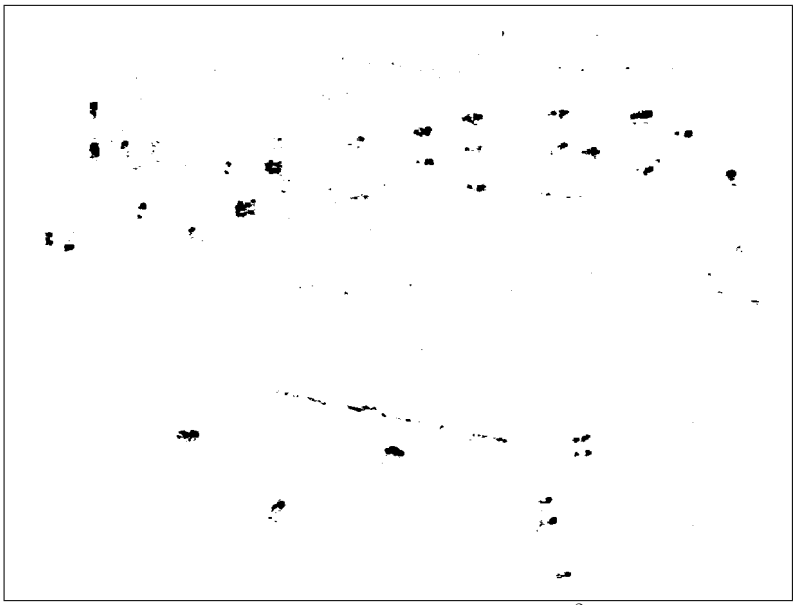
The change detection maps obtained for the case with $W = 5$ and $N = 3$ for the detectors (1.20) - (1.25) are shown in Figures 1.17, 1.18, and 1.19, whereas, the corresponding maps of changes are shown in Figures 1.20, 1.21, and 1.22. Based on them as well as on the results in Table 1.1, it can be observed that detector (1.25) ensures the best performance for $W = 5$, whereas detector (1.20) outperforms the counterparts for $W = 9$. Moreover, for $W = 3$, additional results (not reported in the table for conciseness) highlights that the best performance is obtained exploiting detector (1.25) processing with the pair VV-HV.

Of particular interest are the tests (1.21) and (1.22): from the corresponding detection maps it is easy to recognize that the former identifies the departures, whereas, the latter identifies the arrivals. This behaviour is also clearly visible in the maps shown in Figure 1.21. In fact, for the detector (1.21) the departures presents higher values and the arrivals the lower ones, while for the detector (1.22) the dual behaviour is obtained.

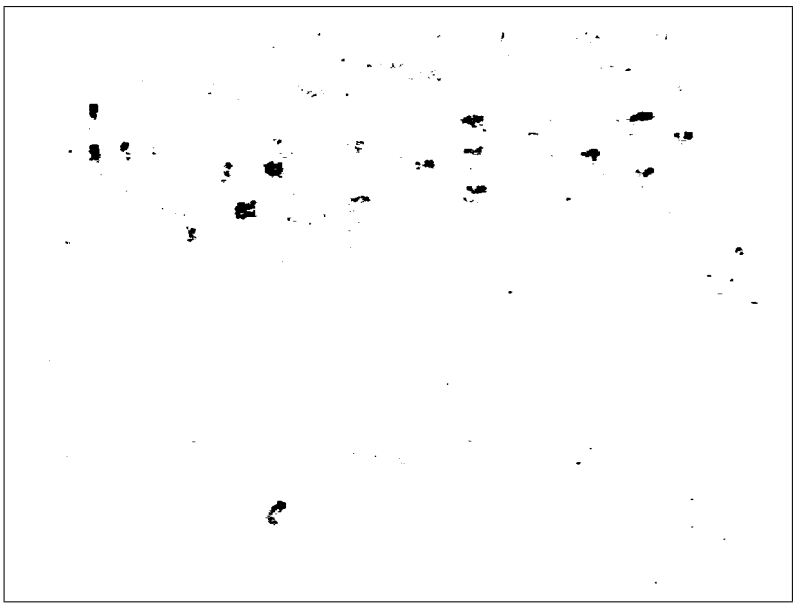
Additionally, although the detector (1.25) is outperformed by detectors (1.20) and (1.23), it allows us to discriminate between the departures and the arrivals based on the argument of the max in (1.25). In particular if the first argument in (1.25) is selected, the change corresponds to a departure, while it corresponds to an arrival if the second argument is selected. By considering the additional information coming from the maximum argument in (1.25) it is thus possible to discriminate the departures from the arrivals detected by the receiver (1.25); in Figure 1.23-b this enhanced detection map is shown together with the ground truth.

Detector			(1.20)	(1.21)	(1.22)	(1.23)	(1.24)	(1.25)	(1.26)
Statistic			$\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$	$\sum_{i=1}^{N} \lambda_i$	$\sum_{i=1}^{N} \frac{1}{\lambda_i}$	$\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$	$\left(\lambda_1 + \frac{1}{\lambda_N}\right)$	$\max\left(\lambda_1\,,\ \frac{1}{\lambda_N}\right)$	$\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$
$W = 5$	$N = 1$	HH	4512	3977	2301	4512	4512	4512	2301
		VV	5190	3941	2485	5190	5190	5190	2759
		HV	5557	4043	2796	5557	5557	5557	2818
	$N = 2$	HH-VV	5372	5129	2337	5103	4877	4980	2300
		HH-HV	6088	4932	3019	5969	5597	5778	2946
		VV-HV	6425	4962	3396	6540	6165	6553	3274
	$N = 3$	HH-VV-HV	6492	5513	2884	5901	5463	5644	2819
$W = 9$	$N = 1$	HH	5671	5426	2765	5671	5671	5671	2797
		VV	7712	5627	3684	7712	7712	7712	5564
		HV	7854	5653	3798	7854	7854	7854	5055
	$N = 2$	HH-VV	6544	6591	2664	5852	5334	5433	2601
		HH-HV	8034	6381	3692	7525	6889	7140	3593
		VV-HV	9009	6773	4580	8994	8355	9005	5086
	$N = 3$	HH-VV-HV	8300	7049	3389	7120	6233	6386	3255

Table 1.1: Number of correct detections in the extended ground truth. All the comparisons have been done using the same P_{fa} level, namely the number of threshold crossings in the complement of the ground truth is exactly the same for all the detectors. The best detector, for the considered dataset and P_{fa} level, is thus the one ensuring the highest number of detections within the ground truth region (number of pixels in the ground truth region where the decision statistic is over the threshold).

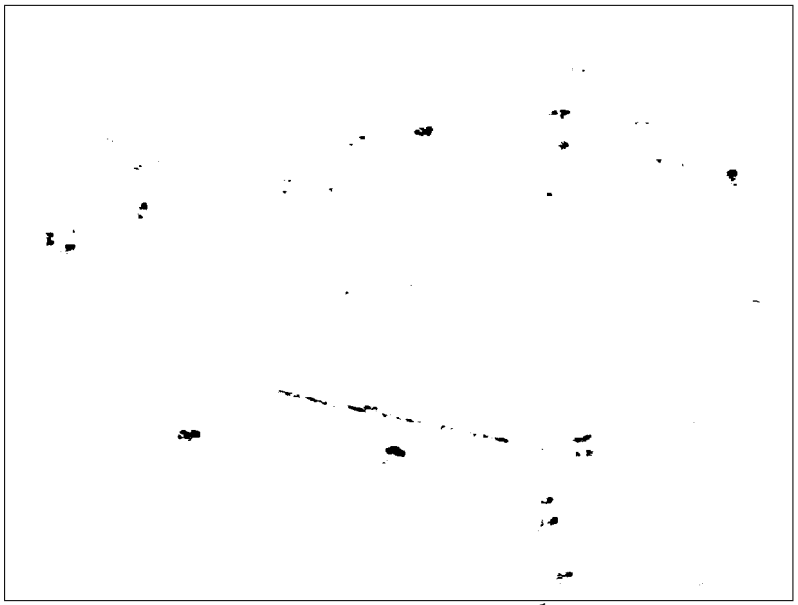


(a) Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$.

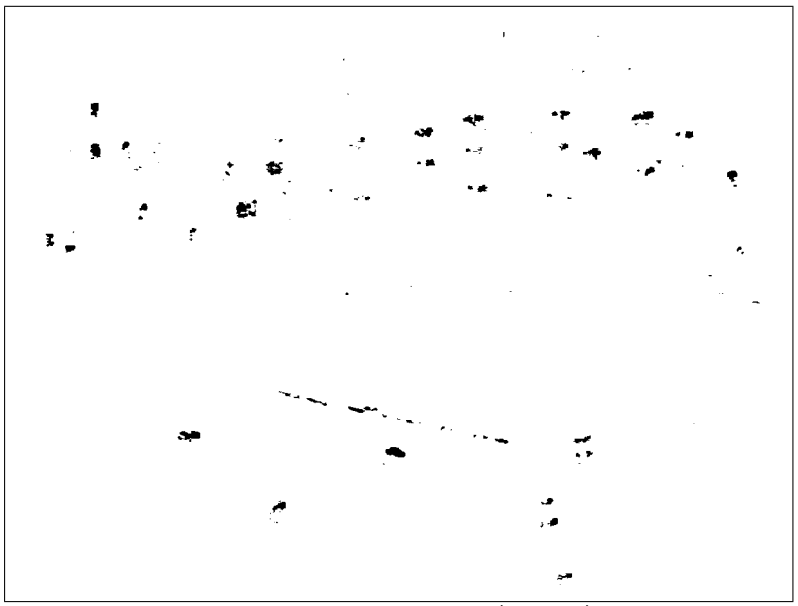


(b) Detector (1.21): $\sum_{i=1}^{N} \lambda_i$.

Figure 1.17: Change Detection maps for $W = 5$ and $N = 3$.

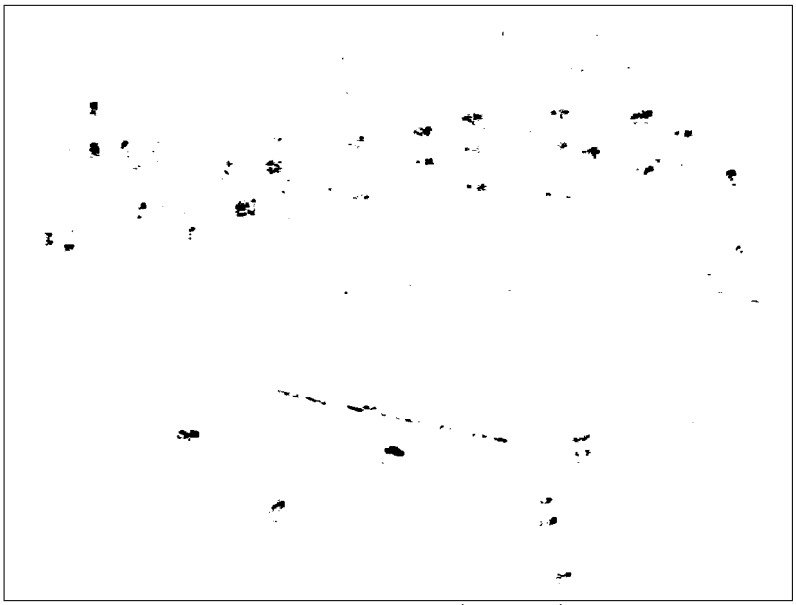


(a) Detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$.

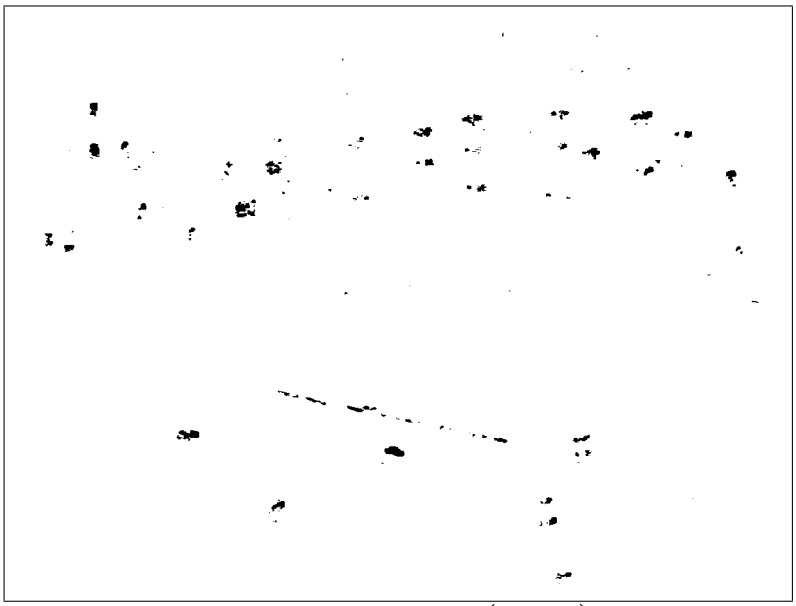


(b) Detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i \right)$.

Figure 1.18: Change Detection maps for $W = 5$ and $N = 3$.



(a) Detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$.



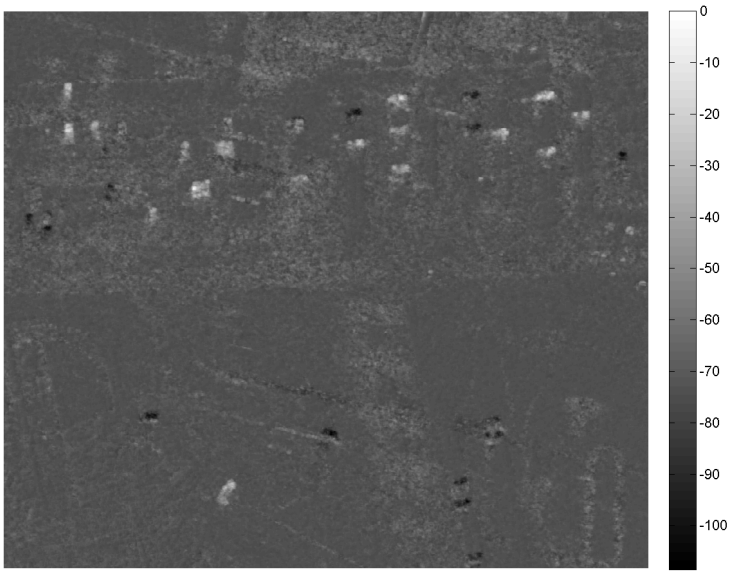
(b) Detector (1.25): $\max\left(\lambda_1\,,\,\frac{1}{\lambda_N}\right)$.

Figure 1.19: Change Detection maps for $W = 5$ and $N = 3$.

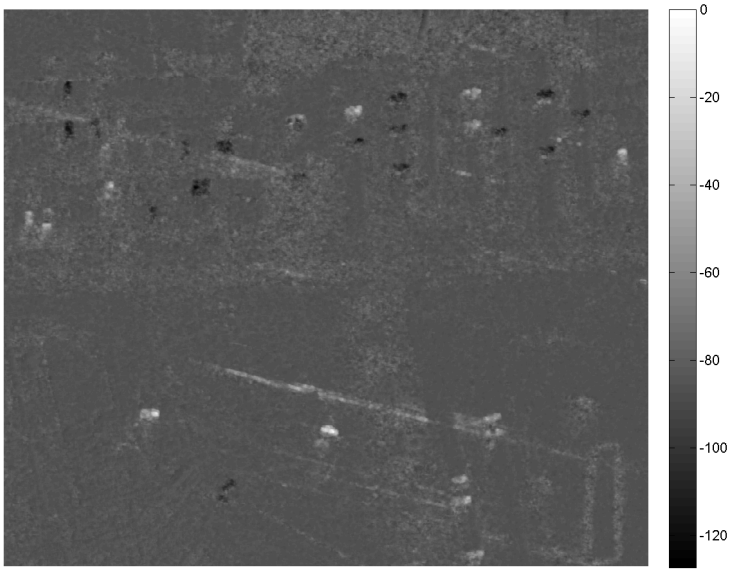
(a) Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$.

(b) Detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i \right)$.

Figure 1.20: Maps of changes in dB for $W = 5$ and $N = 3$.



(a) Detector (1.21): $\sum_{i=1}^{N} \lambda_i$.



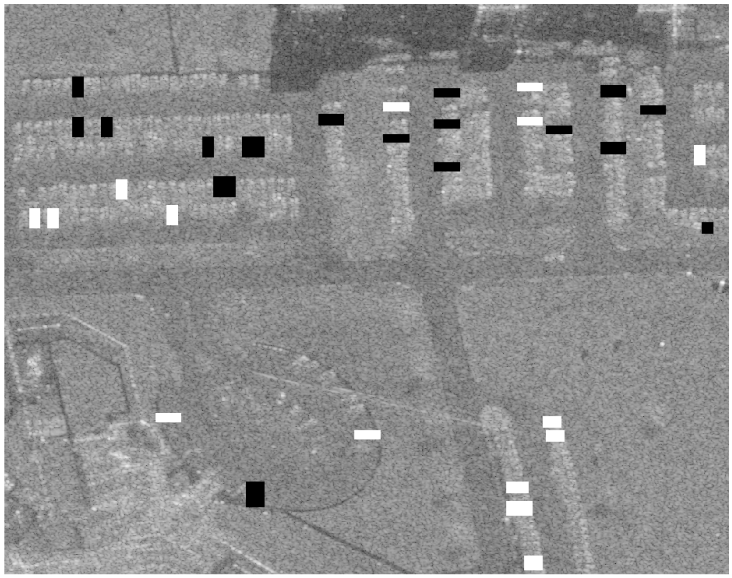
(b) Detector (1.22): $\sum_{i=1}^{N} \frac{1}{\lambda_i}$.

Figure 1.21: Maps of changes in dB for $W = 5$ and $N = 3$.

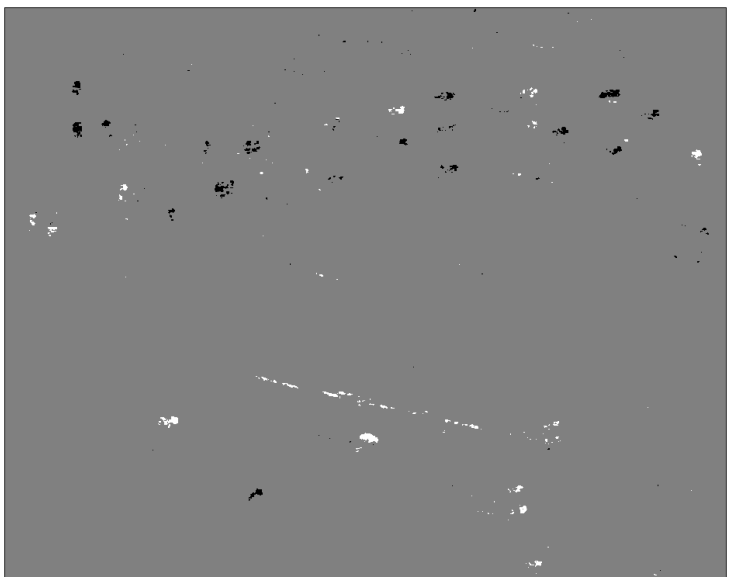
(a) Detector (1.24): $\left(\lambda_1 + \frac{1}{\lambda_N}\right)$.

(b) Detector (1.25): $\max\left(\lambda_1\,,\,\frac{1}{\lambda_N}\right)$.

Figure 1.22: Maps of changes in dB for $W = 5$ and $N = 3$.

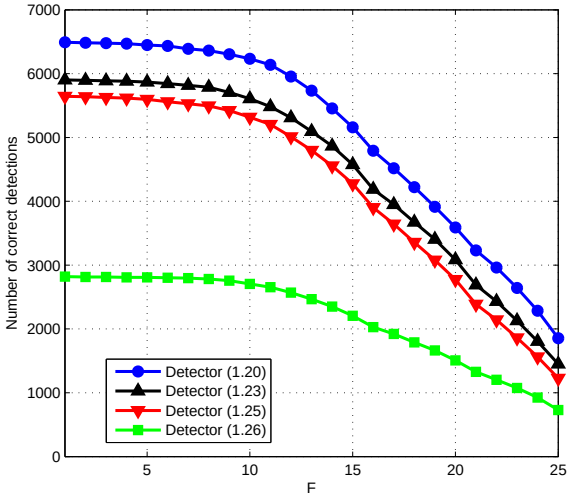


(a) Ground Truth.

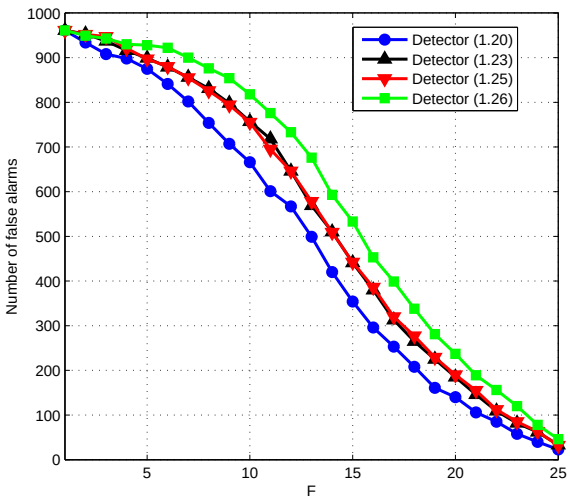


(b) Detector (1.25).

Figure 1.23: Ground truth and enhanced detection map considering $W = 5$ and $N = 3$ for detector (1.25): $\max\left(\lambda_1, \frac{1}{\lambda_N}\right)$. On both images in black and white are shown the departures and the arrivals respectively.



(a) Correct Detections



(b) False Alarms

Figure 1.24: Number of correct detections in the extended ground truth (a) and number of false alarms in the complement of the extended ground truth (b) versus F . Detector (1.20): $\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i}$, detector (1.23): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} + \lambda_i\right)$, detector (1.25): $\max\left(\lambda_1\,,\, \frac{1}{\lambda_N}\right)$, detector (1.26): $\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i}\right)$.

1.5.1 Assessing the Impact of an Aggregation Procedure

In this section are analyzed the effects of an aggregation strategy after single pixel detection so as to eliminate isolated false alarms and confirm true detections, which due to the typical car size and system resolution, appear quite clustered (see the detection maps of Figure 1.17, 1.18, and 1.19). More precisely, a window of size 5×5 slides along the horizontal and the vertical dimensions of the detection maps (which, as already highlighted are binary images: "0" no detection, "1" detection). Then, for a given pixel localized at the center of the moving window and labeled with "1", a detection is associated if the number of "1" in the window is greater that a certain integer "fill parameter" (denoted by F and complying with $1 \leq F \leq 25$).

These kind of logics look like n-of-m aggregation techniques [34, Ch. 3] and can be interpreted as a post-detection binary integration within the reference window. As to the pixels on the image edge (namely those laying on the first/last two rows or columns), no aggregation is performed because they never fall at the center of the moving window. Otherwise stated, it is simply confirmed the "0" or "1" value in the original detection map.

The case of three polarimetric channels is considered and the effect of the fill parameter is studied in Figure 1.24 for detectors (1.20), (1.23), (1.25) and (1.26).

The conducted deleted analysis highlights that P_d is not very sensitive to the fill parameter when it ranges between 5% to 40% of the window size. A similar result was also found in [9] with reference to single channel change detection. Interestingly, there exists a range of values for the parameter F (precisely $1 \leq F \leq 10$) where P_d is almost flat while P_{fa} decreases as F increases. This suggests that a fill parameter value between 8 and 10 can be reasonably used in the aggregation procedure, reducing the number of false alarms while keeping (almost) constant the number of true positives.

1.6 Conclusions

Multi-polarization SAR change detection has been considered in this chapter. The problem has been formulated as a binary hypothesis test and the principle of invariance has been applied to design decision rules exhibiting a special symmetry, which is a sufficient condition to ensure

the CFAR property. A maximal invariant statistic is found and the optimum invariant detector has been computed showing that it is not UMPI and, as a consequence, not practically implementable. Hence, some sub-optimum invariant tests, whose decision statistics are given in terms of the obtained maximal invariant, have been introduced and assessed both on simulated data and on real high resolution SAR images. The conducted analysis has shown that some of them represent a viable mean to deal with the change detection problem. An important result relates to the performance obtained considering the VV-HV polarization. This particular choice (for the considered real dataset) often resulted to provide superior performance compared with all the other single and multi-polarization cases. Moreover, it is also possible to exploit suitable invariant tests for post-detection classification purposes. In particular, from the available real data, to discriminate between car arrivals and departures in the parking area without any additional RCS comparison. Possible future research tracks will consider the extension of the framework relaxing the Gaussian requirement for the data as well as the analysis on other datasets acquired by a different system, possibly at different resolutions (different performance behaviors could be observed on different datasets). Moreover, in order to deal with possible amplitude and phase calibration errors, future investigations will deal with the design of decision rules robust to miscalibration effects. The idea is to achieve this goal by imposing other invariances at the price of some detection losses. Finally, it would be interesting to investigate the joint use of multi-frequency and multi-polarization data to further improve the effect of diversity.

Chapter 2

Unstructured Versus Structured GLRT for Multi-Polarization SAR Change Detection

This chapter deals with coherent (in the sense that both amplitudes and relative phases of the polarimetric returns are used to construct the decision statistic), multi-polarization SAR change detection assuming the availability of reference and test images collected from N multiple polarimetric channels. The problem is formulated as a binary hypothesis testing problem and a special block-diagonal structure for the polarimetric covariance matrix is forced to design a detector based on the Generalized Likelihood Ratio Test (GLRT) criterion. It is shown that the considered decision rule ensures the Constant False Alarm Rate (CFAR) property with respect to the unknown disturbance covariance. At the analysis stage, results on both simulated and real multi-polarization SAR images show the effectiveness of the proposed decision rule and its superiority against the traditional unstructured GLRT in some scenarios of practical interest. Precisely, the chapter is organized as follows. The multi-polarization SAR change detection problem is formulated and the GLRT proposed in [9] is reported in Section 2.1. In Section 2.2 the derivation of the structured GLRT is introduced while, in Section 2.3, the performance is assessed. Finally, in Section 2.4, conclusions are provided.

2.1 Problem Formulation

In the following we suppose the availability of two datacubes $\boldsymbol{X}$ (reference data) and $\boldsymbol{Y}$ (test data) of the same geographic area as considered in Section 1.1.

Herein, we only consider the case of three polarimetric channels and that the available returns from the same pixel are stacked in the specific order HH, VV, and HV. Furthermore, the columns of $\boldsymbol{R}_X$ ($\boldsymbol{R}_Y$) are assumed statistically independent and identically distributed random vectors drawn from a complex circular zero-mean Gaussian distribution with positive definite covariance matrix $\boldsymbol{\Sigma}_X$ ($\boldsymbol{\Sigma}_Y$), complying with the structure introduced in [22], i.e.

$$\boldsymbol{\Sigma}_X \in \Xi \qquad (\boldsymbol{\Sigma}_Y \in \Xi),$$

where

$$\Xi = \left\{ \boldsymbol{\Sigma} \in \mathcal{H}_N^{++} \, : \, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \sigma^2 \end{pmatrix} \right\}, \tag{2.1}$$

With these assumptions, the problem of detecting the presence of possible changes in a rectangular neighbourhood $\mathcal{A}$, with size $K = W_1 \times W_2 \geq N$, of a given pixel can be formulated in terms of the following binary hypothesis test

$$\begin{cases} H_0 : \boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y \\ H_1 : \boldsymbol{\Sigma}_X \neq \boldsymbol{\Sigma}_Y \end{cases} \tag{2.2}$$

where the null hypothesis H_0 of change absence is tested versus the alternative H_1. Exploiting the Gaussian assumption together with the covariance structure (2.1), we can write the joint probability density function (pdf) of $\boldsymbol{R}_X$ and $\boldsymbol{R}_Y$ as[1]

$$\begin{aligned} &f_{\boldsymbol{R}_X,\boldsymbol{R}_Y}(\boldsymbol{R}_X, \boldsymbol{R}_Y | H_1, \boldsymbol{\Sigma}_{X,1}\boldsymbol{\Sigma}_{Y,1}, \sigma^2_{X,1}, \sigma^2_{Y,1}) = \\ &\frac{1}{\pi^{6K} \det^K(\boldsymbol{\Sigma}_{X,1}) \det^K(\boldsymbol{\Sigma}_{Y,1}) \sigma^{2K}_{X,1} \sigma^{2K}_{Y,1}} \times \\ &\exp\left\{ -\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}_{X,1} \boldsymbol{S}_{X,1} + \boldsymbol{\Sigma}^{-1}_{Y,1} \boldsymbol{S}_{Y,1} \right) - \frac{\widehat{\sigma^2}_{X,1}}{\sigma^2_{X,1}} - \frac{\widehat{\sigma^2}_{Y,1}}{\sigma^2_{Y,1}} \right\}, \end{aligned} \tag{2.3}$$

[1] The inverse of a block-diagonal matrix in the class Ξ still belongs to Ξ. This property is used to compute tr $\left(\boldsymbol{\Sigma}_X^{-1} \boldsymbol{S}_X\right)$ and tr $\left(\boldsymbol{\Sigma}_Y^{-1} \boldsymbol{S}_Y\right)$.

and

$$f_{\boldsymbol{R}_X,\boldsymbol{R}_Y}(\boldsymbol{R}_X,\boldsymbol{R}_Y|H_0,\boldsymbol{\Sigma}_{X,1},\sigma^2_{X,1}) = \frac{1}{\pi^{6K}\det^{2K}(\boldsymbol{\Sigma}_{X,1})\sigma^{4K}_{X,1}}\times \exp\left\{-\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}_{X,1}(\boldsymbol{S}_{X,1}+\boldsymbol{S}_{Y,1})\right] - \frac{\widehat{\sigma^2}_{X,1}+\widehat{\sigma^2}_{Y,1}}{\sigma^2_{X,1}}\right\}, \tag{2.4}$$

where $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ are partitioned as

$$\boldsymbol{S}_X = \begin{bmatrix}\boldsymbol{S}_{X,1} & \boldsymbol{S}_{X,2}\\ \boldsymbol{S}^{\dagger}_{X,2} & \widehat{\sigma^2}_{X,1}\end{bmatrix} \qquad \boldsymbol{S}_Y = \begin{bmatrix}\boldsymbol{S}_{Y,1} & \boldsymbol{S}_{Y,2}\\ \boldsymbol{S}^{\dagger}_{Y,2} & \widehat{\sigma^2}_{Y,1}\end{bmatrix}, \tag{2.5}$$

and $\widehat{\sigma^2}_{X,1}$ and $\widehat{\sigma^2}_{Y,1}$ are scalars.
In [9], the GLRT has been devised without considering the special structure (2.1) for $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$. The resulting detector, referred to in the following as unstructured GLRT, is

$$\frac{\det^2(\boldsymbol{S}_X+\boldsymbol{S}_Y)}{\det(\boldsymbol{S}_X)\det(\boldsymbol{S}_Y)} \underset{H_0}{\overset{H_1}{\gtrless}} T_U\,, \tag{2.6}$$

where T_U is the detection threshold set to ensure a given P_{fa} level. In the next section, we exploit the special covariance structure (2.1) induced by polarization diversity and derive the structured GLRT.

2.2 Structured GLRT Design

This approach is equivalent to replacing the unknown parameters in the likelihood ratio with their maximum likelihood estimates, under each hypothesis [31]. Specifically, the structured GLRT is the decision rule (2.7), and after substituting the pdfs defined in (2.3) and (2.4), we get (2.8).

$$\frac{\max\limits_{\boldsymbol{\Sigma}_{X,1},\boldsymbol{\Sigma}_{Y,1},\sigma^2_{X,1},\sigma^2_{Y,1}} f_{\boldsymbol{R}_X,\boldsymbol{R}_Y}(\boldsymbol{R}_X,\boldsymbol{R}_Y|H_1,\boldsymbol{\Sigma}_{X,1}\boldsymbol{\Sigma}_{Y,1},\sigma^2_{X,1},\sigma^2_{Y,1})}{\max\limits_{\boldsymbol{\Sigma}_{X,1},\sigma^2_{X,1}} f_{\boldsymbol{R}_X,\boldsymbol{R}_Y}(\boldsymbol{R}_X,\boldsymbol{R}_Y|H_0,\boldsymbol{\Sigma}_{X,1},\sigma^2_{X,1})} \underset{H_0}{\overset{H_1}{\gtrless}} T_{S,0}\,. \tag{2.7}$$

$$\frac{\max\limits_{\boldsymbol{\Sigma}_{X,1},\boldsymbol{\Sigma}_{Y,1},\sigma^2_{X,1},\sigma^2_{Y,1}} \left\{ \frac{\exp\left[-\mathrm{tr}\left(\boldsymbol{\Sigma}_{X,1}^{-1}\boldsymbol{S}_{X,1}+\boldsymbol{\Sigma}_{Y,1}^{-1}\boldsymbol{S}_{Y,1}\right)-\frac{\widehat{\sigma^2}_{X,1}}{\sigma^2_{X,1}}-\frac{\widehat{\sigma^2}_{Y,1}}{\sigma^2_{Y,1}}\right]}{\pi^{6K}\det^K(\boldsymbol{\Sigma}_{X,1})\det^K(\boldsymbol{\Sigma}_{Y,1})\sigma^{2K}_{X,1}\sigma^{2K}_{Y,1}} \right\}}{\max\limits_{\boldsymbol{\Sigma}_{X,1},\sigma^2_{X,1}} \left\{ \frac{\exp\left[-\mathrm{tr}\left[\boldsymbol{\Sigma}_{X,1}^{-1}(\boldsymbol{S}_{X,1}+\boldsymbol{S}_{Y,1})\right]-\frac{\widehat{\sigma^2}_{X,1}+\widehat{\sigma^2}_{Y,1}}{\sigma^2_{X,1}}\right]}{\pi^{6K}\det^{2K}(\boldsymbol{\Sigma}_{X,1})\sigma^{4K}_{X,1}} \right\}} \underset{H_0}{\overset{H_1}{\gtrless}} T_{S,0}\,. \quad (2.8)$$

Hence, performing the maximizations over the parameters we can recast (2.8) in the equivalent form

$$\frac{\det^{2K}(\boldsymbol{S}_{X,1}+\boldsymbol{S}_{Y,1})}{\det^K(\boldsymbol{S}_{X,1})\det^K(\boldsymbol{S}_{Y,1})} \frac{\left(\widehat{\sigma^2}_{X,1}+\widehat{\sigma^2}_{Y,1}\right)^{2K}}{\left(\widehat{\sigma^2}_{X,1}\widehat{\sigma^2}_{Y,1}\right)^{K}} \underset{H_0}{\overset{H_1}{\gtrless}} T_{S,1}\,, \quad (2.9)$$

with $T_{S,1}$ a modified version of $T_{S,0}$. Finally, after a monotonic transformation, we get the following equivalent form of the GLRT

$$\frac{\det^{2}(\boldsymbol{S}_{X,1}+\boldsymbol{S}_{Y,1})}{\det(\boldsymbol{S}_{X,1})\det(\boldsymbol{S}_{Y,1})} \frac{\left(\widehat{\sigma^2}_{X,1}+\widehat{\sigma^2}_{Y,1}\right)^{2}}{\widehat{\sigma^2}_{X,1}\widehat{\sigma^2}_{Y,1}} \underset{H_0}{\overset{H_1}{\gtrless}} T_{S}\,, \quad (2.10)$$

with T_S the modified detection threshold. It can be proved that (2.10), ensures the CFAR property with respect to both $\boldsymbol{\Sigma}_{X,1}$ and $\sigma^2_{X,1}$ (see Appendix B which also provides a stochastic representation for the fast simulation of the detector). Otherwise stated, the detection threshold ensuring a given False Alarm Rate (FAR) can be set independent of the two aforementioned parameters.

2.3 Performance Analysis

This section presents the performance analysis for the proposed decision rule for both simulated and real data.

2.3.1 Performance Analysis on Simulated Data

This sub-section presents the performance analysis via computer simulated data of the detectors introduced in Sections 2.1 and 2.2. In particular, the standard ROCs are computed for the unstructured and

structured GLRTs and compared with the benchmark performance of the optimum Neyman-Pearson detector. In order to set the detection threshold, Monte Carlo simulations are used assuming $100/P_{fa}$ independent runs. Additionally, 10^5 independent trials are exploited to estimate P_d. As in [9] the theoretical covariance matrices considered to estimate the P_d are:

$$\boldsymbol{\Sigma}_X = \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 0.2 \end{pmatrix} \qquad \boldsymbol{\Sigma}_Y = 2\boldsymbol{\Sigma}_X,$$

while $\boldsymbol{\Sigma}_Y = \boldsymbol{\Sigma}_X$ was considered to estimate the P_{fa}.

The optimum receiver assumes that the actual covariance matrices are known, and can be expressed as:

$$\text{tr}\left[\left(\boldsymbol{\Sigma}_X^{-1} - \boldsymbol{\Sigma}_Y^{-1}\right)\boldsymbol{S}_Y\right] \underset{H_0}{\overset{H_1}{\gtrless}} T, \tag{2.11}$$

which resorting to the special structure of $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ leads to

$$\text{tr}\left[\left(\boldsymbol{\Sigma}_{X,1}^{-1} - \boldsymbol{\Sigma}_{Y,1}^{-1}\right)\boldsymbol{S}_{Y,1}\right] + \left(\frac{1}{\sigma_{X,1}^2} - \frac{1}{\sigma_{Y,1}^2}\right)\hat{\sigma}_{Y,1}^2 \underset{H_0}{\overset{H_1}{\gtrless}} T. \tag{2.12}$$

The obtained ROCs for the cases of $W = 3, 5$ and 7 are shown in Figures 2.1, 2.2 and 2.3 respectively. In all cases the structured GLRT outperforms the unstructured one, while the optimum receiver provides the benchmark performance. For example for the case of $W = 5$ with a P_{fa} of 10^{-4} the P_d assumes value of 0.1386 for detector (2.6) while it is 0.2822 for detector (2.10) and 0.9913 for the detector (2.12). Moreover it is worthwhile to note that for all the detectors the P_d improves as W increases for a given P_{fa}. This effect is principally due to the more accurate estimation of the covariance matrices which exploits more homogeneous data.

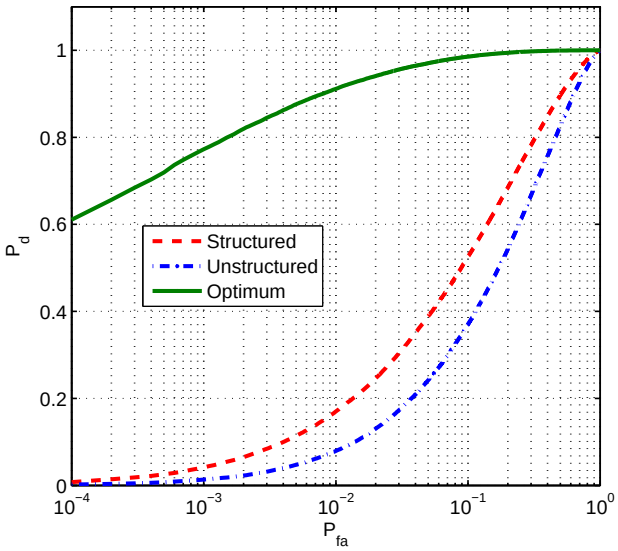


Figure 2.1: P_d versus P_{fa} for $W = 3$.

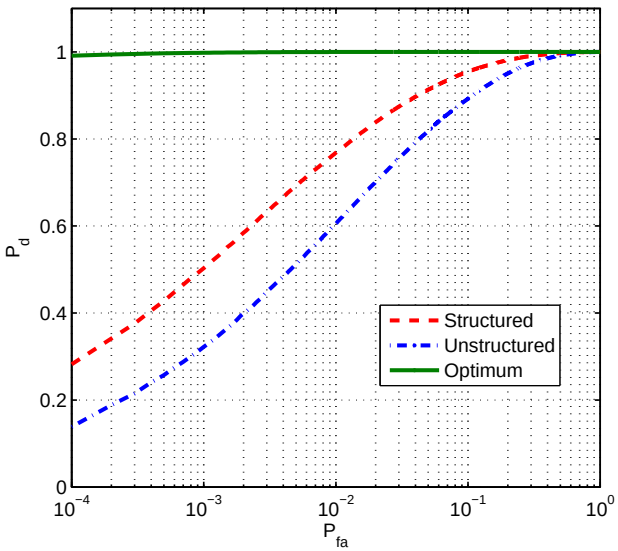


Figure 2.2: P_d versus P_{fa} for $W = 5$.

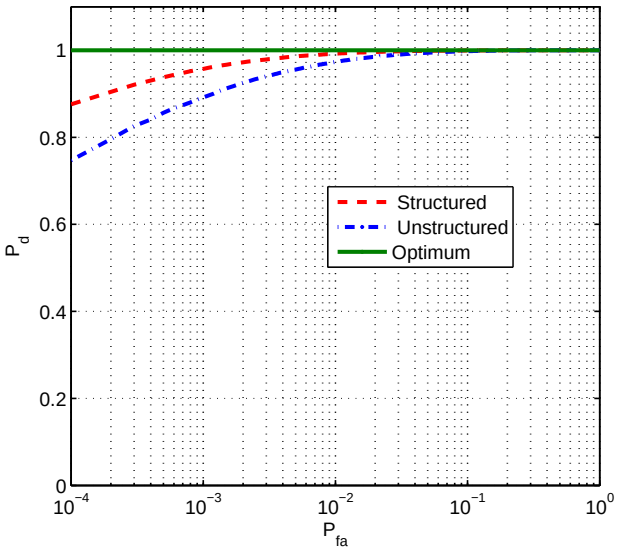


Figure 2.3: P_d versus P_{fa} for $W = 7$.

2.3.2 Testing on Real Data

In this subsection, the performance analysis on the real data described in Chapter 1 is presented. In particular, the thresholds are set to ensure $P_{fa} = 10^{-3}$ in the complement of the extended ground truth area, namely, in the region where no changes occur (there are no true positives). This means that, for each detector, after computing the decision statistics (for each pixel belonging to the complement of the extended ground truth), the threshold has been selected in order that

$$10^{-3} \times \text{total number of available statistics (trials)}\,,$$

are greater than the threshold. This ensures that all the comparisons refer to the same P_{fa} level, namely the number of threshold crossings in the complement of the extended ground truth is exactly the same for all the analysed detectors. In Table 2.1 the number of correct changes detected using receivers (2.6) and (2.10) for the cases with $W = 3$, $W = 5$, and $W = 7$ are reported.

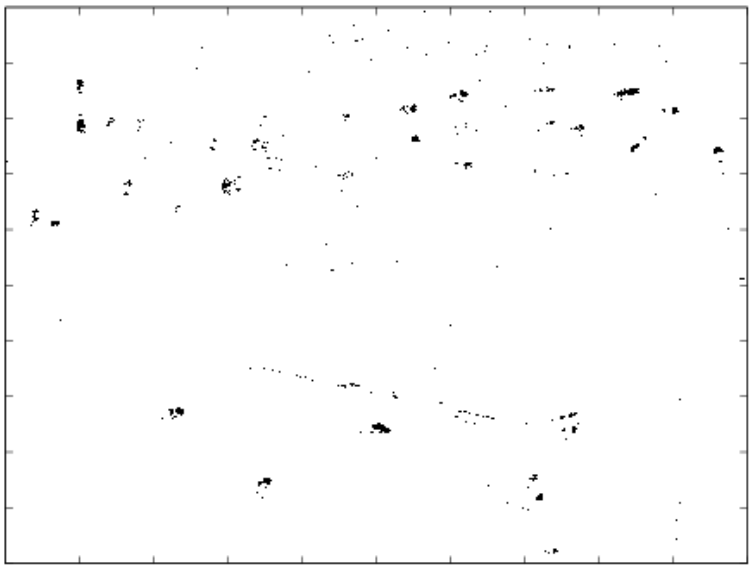
	W		
Detector	3	5	7
Unstructured GLRT (2.6)	3802	6492	7533
Structured GLRT (2.10)	4949	6655	7387

Table 2.1: Number of correct detections for $W = 3$, 5 and 7.

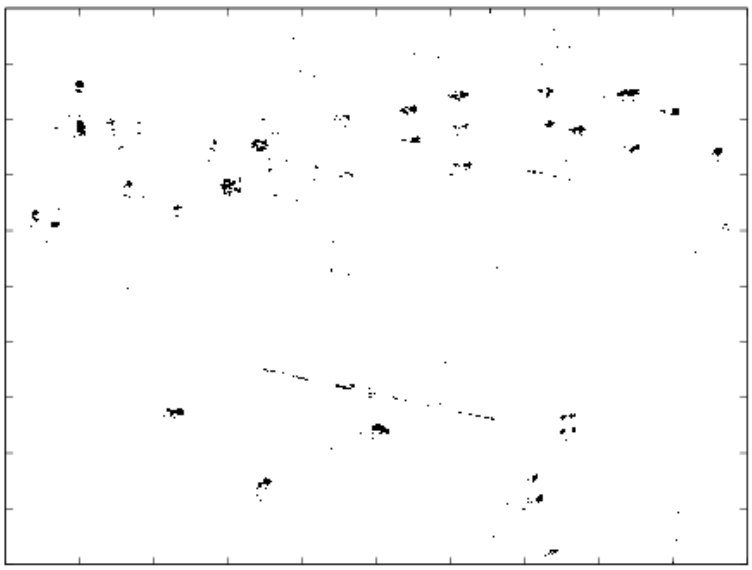
From Table 2.1 it is clear that the structured GLRT outperforms the unstructured GLRT for the smaller window sizes ($W = 3$ and 5) whereas the unstructured GLRT outperforms the structured GLRT for the larger window size of $W = 7$ when it is able to detect more changes in the image. This last result can be justified in terms of a covariance model mismatch in the sense that the off-diagonal entries $(1, 3)$ and $(3, 1)$ of the polarimetric covariance matrix which in the theoretical model have been set to zero might not be exactly zero in reality (even if very close to that value). Additionally there might be some other deviations from the theoretical model for instance due to environmental non-homogeneities. Last results presented are example of detection maps for detectors (2.6) and (2.10) for the case of $W = 3$ in Figure 2.4-a and in Figure 2.4-b respectively. From the detection maps the higher number of detection achievable with the detector (2.10) is appreciable.

2.4 Conclusions

The block-diagonal structure for the polarimetric covariance matrix is exploited in this chapter to derive a decision rule based on the GLRT criterion. The proposed approach ensures the CFAR property with respect to the unknown polarimetric covariance. The structured GLRT detector has been compared with the unstructured GLRT, with analysis on both simulated and real full-polarimetric SAR data. The performance analysis confirmed that a structured approach can provide increased performance with particular benefits when a small amount of homogeneous data is available. Future work will concentrate on the performance analysis on other datasets, the study of model sensitivity to mismatches, and the development of an invariant framework accounting at the maximal invariant design stage of the block-diagonal structure of the polarimetric covariance matrix.



(a) Detector (2.6).



(b) Detector (2.10).

Figure 2.4: Detection maps for $W = 3$.

Chapter 3

Forcing Scale-Invariance in Multi-Polarization SAR Change Detection

This chapter considers the problem of coherent multi-polarization SAR change detection starting from the availability of image pairs exhibiting possible power mismatches/miscalibrations. The principle of invariance is used to characterize the class of scale-invariant decision rules which are insensitive to power mismatches and ensure the Constant False Alarm Rate (CFAR) property. A maximal invariant statistic is derived together with the induced maximal invariant in the parameter space which significantly compress the data/parameters domain. A Generalized Likelihood Ratio Test (GLRT) is synthesized both for the cases of two- and three-polarimetric channels. The performance of the proposed scale-invariant structures is compared to those from two non-invariant counterparts using both simulated and real radar data. The results highlight the robustness of the proposed method and the performance tradeoff involved. Precisely, in Section 3.1, we deal with the formulation of the multi-polarization SAR change detection problem. In Section 3.2, the maximal invariant for the scale-invariant SAR change detection problem is established. The design of the GLRT and other scale-invariant receivers is presented in Section 3.3, whereas, in Section 3.4, the performance of the introduced invariant tests is shown on both simulated and real multi-polarization SAR images. Finally, conclusions are provided in Section 3.5.

3.1 Problem Formulation

In the following, starting from the framework proposed in Chapter 1, we focus on the problem of change detection when the image pairs involved in the change detection procedure can exhibit possible power mismatches/miscalibrations. To this end, the change detection problem in the region $\mathcal{A}$ can be formulated in terms of the following binary hypothesis test

$$\begin{cases} H_0 : \boldsymbol{\Sigma}_X = \gamma \boldsymbol{\Sigma}_Y \\ H_1 : \boldsymbol{\Sigma}_X \neq \gamma \boldsymbol{\Sigma}_Y \end{cases} \tag{3.1}$$

where the null hypothesis H_0 of change absence is tested versus the alternative H_1 and the parameter $\gamma > 0$ models possible received power variations between two different acquisitions from the same scene (mainly due to misalignment of the flight paths as well as channel propagation effects).

Exploiting the Gaussian assumption, we can write the joint probability density function (pdf) of $\boldsymbol{R}_X$ and $\boldsymbol{R}_Y$ as

$$\begin{aligned} &f_{\boldsymbol{R}_X,\boldsymbol{R}_Y}(\boldsymbol{R}_X, \boldsymbol{R}_Y | H_1, \boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y) = \\ &\qquad \frac{1}{\pi^{2NK} \det(\boldsymbol{\Sigma}_X \boldsymbol{\Sigma}_Y)^K} \exp\left\{-\text{tr}\left(\boldsymbol{\Sigma}_X^{-1}\boldsymbol{S}_X + \boldsymbol{\Sigma}_Y^{-1}\boldsymbol{S}_Y\right)\right\} . \end{aligned} \tag{3.2}$$

Using the Fisher-Neyman factorization theorem [25], we can claim that a sufficient statistic for (3.1) is represented by the two sample Grammian matrices $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ which are statistically independent and follow a complex Wishart distribution, i.e. [26]

$$\begin{aligned} &f_{\boldsymbol{S}_X}(\boldsymbol{S}_X | H_1, \boldsymbol{\Sigma}_X) = \\ &\qquad \frac{c_W}{\det(\boldsymbol{\Sigma}_X)^K} \exp\left\{-\text{tr}\left(\boldsymbol{\Sigma}_X^{-1}\boldsymbol{S}_X\right)\right\} \det(\boldsymbol{S}_X)^{K-N} , \quad \boldsymbol{S}_X \succ \boldsymbol{0} \end{aligned} \tag{3.3}$$

$$\begin{aligned} &f_{\boldsymbol{S}_Y}(\boldsymbol{S}_Y | H_1, \boldsymbol{\Sigma}_Y) = \\ &\qquad \frac{c_W}{\det(\boldsymbol{\Sigma}_Y)^K} \exp - \left\{\text{tr}\left(\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{S}_Y\right)\right\} \det(\boldsymbol{S}_Y)^{K-N} , \quad \boldsymbol{S}_Y \succ \boldsymbol{0} \end{aligned} \tag{3.4}$$

with c_W a proper normalization constant. From the sufficient statistic we can evaluate the optimum Neyman-Pearson (NP) detector as the

Likelihood Ratio Test (LRT), which, after standard algebra and statistical equivalences, can be recast as

$$\mathrm{tr}\left[\left(\frac{\boldsymbol{\Sigma}_Y^{-1}}{\gamma}-\boldsymbol{\Sigma}_X^{-1}\right)\boldsymbol{S}_X\right]\underset{H_0}{\overset{H_1}{\gtrless}}T_0\,, \tag{3.5}$$

where T_0 is the detection threshold. Evidently, test (3.5) is not Uniformly Most Powerful (UMP) and, consequently, it is not practically implementable because it requires the knowledge of $\boldsymbol{\Sigma}_X$, γ, and $\boldsymbol{\Sigma}_Y$ which, in realistic applications, are usually unknown.

3.2 Data Reduction and Invariance Issues

Both hypotheses under test are composite or, otherwise stated, H_0 and H_1 are equivalent to a partition of the parameter space Θ into the two disjoint sets

$$\begin{aligned}\Theta_0&=\left\{\boldsymbol{\Sigma}_X=\gamma\boldsymbol{\Sigma}_Y,\ (\boldsymbol{\Sigma}_X,\boldsymbol{\Sigma}_Y,\gamma)\in\mathcal{H}_N^{++}\times\mathcal{H}_N^{++}\times\mathcal{R}^{++}\right\}\\ \Theta_1&=\left\{\boldsymbol{\Sigma}_X\neq\gamma\boldsymbol{\Sigma}_Y,\ (\boldsymbol{\Sigma}_X,\boldsymbol{\Sigma}_Y,\gamma)\in\mathcal{H}_N^{++}\times\mathcal{H}_N^{++}\times\mathcal{R}^{++}\right\}\end{aligned}\,. \tag{3.6}$$

This formulation indicates that the individual values of the nuisance parameters are irrelevant: one must only discern to which hypothesis they belong to, namely whether the covariances are proportional or not. This observation highlights that we can cluster the data considering transformations that leave unaltered:

a. the two composite hypotheses, namely the partition of the parameter space;

b. the families of distributions under the two hypotheses.

This can be achieved through the *Principle of Invariance* [16] by which we look for transformations that preserve the formal structure of the hypothesis testing problem and, then, for decision rules invariant to them while also acting as a data reduction technique (i.e., leading to an observation space of significantly lower dimensionality than the original one).

It is not difficult to prove that our testing problem is invariant under the group of transformations G acting on the sufficient statistic as[1]:

$$G = \left\{ g \,:\, \boldsymbol{S}_X \to \boldsymbol{B}\boldsymbol{S}_X\boldsymbol{B}^\dagger \,,\, \boldsymbol{S}_Y \to a\boldsymbol{B}\boldsymbol{S}_Y\boldsymbol{B}^\dagger \,,\, \boldsymbol{B} \in \mathcal{GL}(N) \,,\, a \in \mathcal{R}^{++} \right\} . \tag{3.7}$$

In fact, the families of distributions are preserved because if $\boldsymbol{S}_X$ ($\boldsymbol{S}_Y$) is Wishart distributed then $\boldsymbol{B}\boldsymbol{S}_X\boldsymbol{B}^\dagger$ ($\boldsymbol{B}\boldsymbol{S}_Y\boldsymbol{B}^\dagger$) is also Wishart with the same scalar parameters and matrix parameter $\boldsymbol{B}^\dagger\boldsymbol{\Sigma}_X\boldsymbol{B}$ ($a\boldsymbol{B}^\dagger\boldsymbol{\Sigma}_Y\boldsymbol{B}$), where $\boldsymbol{B} \in \mathcal{GL}(N)$ and $a > 0$. Moreover, the original partition of the parameter space is left unaltered since if $\boldsymbol{\Sigma}_X \neq \gamma\boldsymbol{\Sigma}_Y$ then $\boldsymbol{B}\boldsymbol{\Sigma}_X\boldsymbol{B}^\dagger \neq a\gamma\boldsymbol{B}\boldsymbol{\Sigma}_Y\boldsymbol{B}^\dagger$ and if $\boldsymbol{\Sigma}_X = \gamma\boldsymbol{\Sigma}_Y$ then $\boldsymbol{B}\boldsymbol{\Sigma}_X\boldsymbol{B}^\dagger = a\gamma_1\boldsymbol{B}\boldsymbol{\Sigma}_Y\boldsymbol{B}^\dagger$.

3.2.1 Maximal Invariant Design

The invariance property induces a partition of the data space into orbits (or equivalence classes) where, over each orbit, every point is related to every other through a transformation which is member of the group G. Any statistic that identifies different orbits in a one-to-one way significantly reduces the total amount of data necessary for solving the hypothesis testing problem and constitutes the compressed data set to be used in the design of any invariant detector. This kind of statistics are called maximal invariant since they are constant over each orbit (invariance) while they assume different values on different orbits (maximality).

Formally, a statistic $\mathbf{T}(\boldsymbol{S}_X, \boldsymbol{S}_Y)$ is said to be a maximal invariant with respect to the group of transformations G if and only if

- Invariance:
  $\mathbf{T}(\boldsymbol{S}_X, \boldsymbol{S}_Y) = \mathbf{T}[g(\boldsymbol{S}_X, \boldsymbol{S}_Y)]$, $\forall g \in G$.
- Maximality:
  $\mathbf{T}(\boldsymbol{S}_{X_1}, \boldsymbol{S}_{Y_1}) = \mathbf{T}(\boldsymbol{S}_{X_2}, \boldsymbol{S}_{Y_2})$ implies that $\exists\, g \in G$ such that

$$(\boldsymbol{S}_{X_2}, \boldsymbol{S}_{Y_2}) = g(\boldsymbol{S}_{X_1}, \boldsymbol{S}_{Y_1}) \,.$$

Notice that there are many maximal invariant statistics, but they are equivalent in that they yield statistically equivalent detectors. Moreover, all invariant tests can be expressed as a function of the maximal invariant

[1] The proof that (3.7) is a group is given in Appendix C.

statistic [15, 27], which for the problem of interest is provided by the following

Proposition 1: A maximal invariant statistic for problem (3.1) with respect to the group of transformations (3.7) is the $(N-1)$-dimensional vector

$$\left(\frac{\lambda_2}{\lambda_1}, \frac{\lambda_3}{\lambda_1} \dots \frac{\lambda_N}{\lambda_1}\right), \tag{3.8}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ are the eigenvalues of $\boldsymbol{S}_X \boldsymbol{S}_Y^{-1}$.

Proof. The transformation group G can be obtained as composition of the sub-groups

$$D = \left\{ d : \boldsymbol{S}_X \to \boldsymbol{B}\boldsymbol{S}_X\boldsymbol{B}^\dagger, \quad \boldsymbol{S}_Y \to \boldsymbol{B}\boldsymbol{S}_Y\boldsymbol{B}^\dagger, \quad \boldsymbol{B} \in \mathcal{GL}(N) \right\} \tag{3.9}$$

and

$$E = \left\{ e : \boldsymbol{S}_X \to \boldsymbol{S}_X, \quad \boldsymbol{S}_Y \to a\boldsymbol{S}_Y, \quad a \in \mathcal{R}^{++} \right\}. \tag{3.10}$$

As a consequence, the maximal invariant can be obtained in two steps, each corresponding to a sub-group of G [16]. To this end, a maximal invariant under the group D has been derived in Chapter 1 and is given by the eigenvalues $\lambda_1, \ldots, \lambda_N$ of $\boldsymbol{S}_X \boldsymbol{S}_Y^{-1}$. Additionally, $\forall e \in E$, namely $\forall a > 0$, if the eigenvalues of $\boldsymbol{S}_{X,1}\boldsymbol{S}_{Y,1}^{-1}$ are equal to those of $\boldsymbol{S}_{X,2}\boldsymbol{S}_{Y,2}^{-1}$, then the eigenvalues of $a^{-1}\boldsymbol{S}_{X,1}\boldsymbol{S}_{Y,1}^{-1}$ coincide with those of $a^{-1}\boldsymbol{S}_{X,2}\boldsymbol{S}_{Y,2}^{-1}$. Now, define the group $E^\star$ of scale change acting on $\boldsymbol{y} = (y_1, \ldots, y_N)^T \in (\mathcal{R}^{++})^N$ as

$$E^\star = \left\{ e^\star : \boldsymbol{y} \to \frac{1}{a}\boldsymbol{y}, \quad a \in \mathcal{R}^{++} \right\}, \tag{3.11}$$

and observe that a maximal invariant with respect to this last group $E^\star$ is

$$\frac{y_2}{y_1}, \ldots, \frac{y_N}{y_1}. \tag{3.12}$$

Hence, [16, Prop. 2, p. 288] implies that

$$\frac{\lambda_2}{\lambda_1}, \ldots, \frac{\lambda_N}{\lambda_1}, \tag{3.13}$$

is a maximal invariant under G.

As desired, the principle of invariance produces a significant data reduction: the maximal invariant statistic is a real N-dimensional vector

whereas the original sufficient statistic is composed by the two $N \times N$ Grammian matrices $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$. A nice physical/intuitive meaning of the eigenvalues of $\boldsymbol{S}_X \boldsymbol{S}_Y^{-1}$, which are involved in the computation of the maximal invariant, stems from the observation that they are solutions to suitable optimization problems involving the generalized Rayleigh quotient (i.e. extrema of the backscattering ratio varying the scattering mechanism [35]).

3.2.2 Induced Maximal Invariant Design

The data transformation induces a parameter transformation which leaves the two composite hypotheses unaltered. In other words, by the principle of invariance, one partitions also the parameter space into orbits and, usually, deals with a reduced set of parameters. The relevant parameters are embodied into any *induced maximal invariant*, namely any function of the parameters that is constant over each orbit of the parameter space (invariance) but assumes different values over different orbits (maximality).

For the case at hand, an induced maximal invariant is composed of $\left(\frac{\delta_2}{\delta_1}, \ldots, \frac{\delta_N}{\delta_1}\right)$, where $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_N]^T$, $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_N$, are the eigenvalues of the matrix:

$$\boldsymbol{\Sigma}_X \boldsymbol{\Sigma}_Y^{-1} . \tag{3.14}$$

The proof of this claim can be done following the same steps as in the proof of Proposition 1 and it has been omitted for the sake of compactness. The physical interpretation of the induced maximal invariant components follows from the observation that they are related to the spread among the backscattering ratios associated with two different scattering mechanisms.

The previous observation highlights that the principle of invariance also yields a significant reduction of the number of the parameters: in fact, the induced maximal invariant is an $(N-1)$-dimensional vector while the original parameter space was composed of the two covariance matrices $\boldsymbol{\Sigma}_X$, $\boldsymbol{\Sigma}_Y$ and γ.

We explicitly observe that in the reduced parameter space the partition corresponding to the two composite hypotheses of the test (3.1) is $\Xi_0 = \{\mathbf{1}_{N-1}\}$, relative to $\boldsymbol{\Sigma}_X = \gamma \boldsymbol{\Sigma}_Y$, and $\Xi_1 = \overline{\{\mathbf{1}_{N-1}\}}$, relative to $\boldsymbol{\Sigma}_X \neq \gamma \boldsymbol{\Sigma}_Y$, where $\overline{\{\mathbf{1}_{N-1}\}}$ is the set of the $(N-1)$-dimensional

column vectors with positive elements and at least one entry different from 1. The structure of Ξ_0, which now corresponds to a simple H_0 hypothesis, clearly shows that all invariant receivers that process a maximal invariant statistic through a transformation independent of $(\omega_1, \ldots, \omega_{N-1}) = \left(\frac{\delta_2}{\delta_1}, \ldots, \frac{\delta_N}{\delta_1}\right)$, achieve the CFAR property with respect both γ and $\boldsymbol{\Sigma}_X$.

3.3 GLRT Derivation

This section is devoted to the derivations of the GLRT detector for the considered problem. Precisely, the following decision rule is considered

$$\frac{\max\limits_{\boldsymbol{\Sigma}_X}\max\limits_{\boldsymbol{\Sigma}_Y}\dfrac{\exp\left[-\mathrm{tr}\,\left(\boldsymbol{\Sigma}_X^{-1}\boldsymbol{S}_X+\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{S}_Y\right)\right]}{(\pi)^{2NK}\det^K(\boldsymbol{\Sigma}_X)\det^K(\boldsymbol{\Sigma}_Y)}}{\max\limits_{\gamma>0}\max\limits_{\boldsymbol{\Sigma}_Y}\dfrac{\exp\left[-\mathrm{tr}\left(\boldsymbol{\Sigma}_Y^{-1}\left(\dfrac{\boldsymbol{S}_X}{\gamma}+\boldsymbol{S}_Y\right)\right)\right]}{(\pi)^{2NK}\gamma^{NK}\det^{2K}(\boldsymbol{\Sigma}_Y)}} \underset{H_0}{\overset{H_1}{\gtrless}} T\,, \tag{3.15}$$

which, after the optimizations over $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ at the numerator and over $\boldsymbol{\Sigma}_Y$ at the denominator can be recast (after some algebra and statistical equivalences) as,

$$\frac{\min\limits_{\gamma>0}\left[\gamma^N\det^2\left(\dfrac{\boldsymbol{S}_X}{\gamma}+\boldsymbol{S}_Y\right)\right]}{\det(\boldsymbol{S}_X)\det(\boldsymbol{S}_Y)} \underset{H_0}{\overset{H_1}{\gtrless}} T_1\,, \tag{3.16}$$

or equivalently as

$$\frac{\min\limits_{\gamma>0}\left[\gamma^N\det^2\left(\dfrac{\boldsymbol{S}_Y^{-\frac{1}{2}}\boldsymbol{S}_X\boldsymbol{S}_Y^{-\frac{1}{2}}}{\gamma}+\boldsymbol{I}\right)\right]}{\det(\boldsymbol{S}_Y^{-\frac{1}{2}}\boldsymbol{S}_X\boldsymbol{S}_Y^{-\frac{1}{2}})} \underset{H_0}{\overset{H_1}{\gtrless}} T_1, \tag{3.17}$$

where T_1 is a suitable modification of the original threshold, T, in (3.15). In order to proceed further, we have to distinguish between the cases of two- ($N = 2$) and three- ($N = 3$) polarimetric channels.

3.3.1 Case $N = 2$

Forcing $N = 2$ in (3.16) yields

$$\frac{\min\limits_{\gamma>0}\left[\gamma^2\left(\frac{\lambda_1}{\gamma}+1\right)^2\left(\frac{\lambda_2}{\gamma}+1\right)^2\right]}{\lambda_1\lambda_2} \underset{H_0}{\overset{H_1}{\gtrless}} T_1\,. \tag{3.18}$$

It is now necessary to compute

$$\min_{\gamma>0}\left[\frac{1}{\gamma}\left(\lambda_1\lambda_2+\gamma^2+(\lambda_1+\lambda_2)\gamma\right)\right]^2. \tag{3.19}$$

Standard arguments on optimization of univariate functions provide the optimal point $\gamma_{opt,2} = \sqrt{\lambda_1\lambda_2}$. As a consequence, the GLRT becomes

$$\left(\sqrt{\frac{\lambda_1}{\lambda_2}}+1\right)^2\left(\sqrt{\frac{\lambda_2}{\lambda_1}}+1\right)^2 \underset{H_0}{\overset{H_1}{\gtrless}} T_1\,. \tag{3.20}$$

Observing that the left hand side of (3.20) is a monotone increasing function of $\sqrt{\frac{\lambda_1}{\lambda_2}}$ for $\sqrt{\frac{\lambda_1}{\lambda_2}} \in [1, +\infty[$, the GLRT (3.20) turns out equivalent to

$$\frac{\lambda_1}{\lambda_2} \underset{H_0}{\overset{H_1}{\gtrless}} T_2\,, \tag{3.21}$$

with T_2 the modified threshold. Two important comments are now in order. First, test (3.21) is equivalent to comparing the condition number of the matrix $\boldsymbol{S}_Y^{-\frac{1}{2}}\boldsymbol{S}_X\boldsymbol{S}_Y^{-\frac{1}{2}}$ with a detection threshold to establish the presence of changes in the considered scene. Second, the GLRT statistic is a maximal invariant.

3.3.2 Case $N = 3$

Forcing $N = 3$ in (3.16) yields

$$\frac{\min\limits_{\gamma>0}\left[\gamma^3\left(\frac{\lambda_1}{\gamma}+1\right)^2\left(\frac{\lambda_2}{\gamma}+1\right)^2\left(\frac{\lambda_3}{\gamma}+1\right)^2\right]}{\lambda_1\lambda_2\lambda_3} \underset{H_0}{\overset{H_1}{\gtrless}} T\,. \tag{3.22}$$

It is thus necessary to solve the optimization problem

$$\min_{\gamma>0}\left[\gamma^{\frac{3}{2}}\left(\frac{\lambda_1}{\gamma}+1\right)\left(\frac{\lambda_2}{\gamma}+1\right)\left(\frac{\lambda_3}{\gamma}+1\right)\right]^2=\min_{\gamma>0} f_3(\gamma)\,. \tag{3.23}$$

Since

$$\lim_{\gamma\to+\infty} f_3(\gamma)=+\infty\,,\qquad \lim_{\gamma\to 0} f_3(\gamma)=+\infty\,, \tag{3.24}$$

and $f_3(\gamma)$ is continuous in $]0\,,+\infty[$, the minimum is achieved in correspondence of $\gamma_{opt}\in]0\,,+\infty[$. Moreover, finding the optimum value of $f_3(\gamma)$ is equivalent to minimizing its logarithm, i.e.

$$\log f_3(\gamma)=\frac{3}{2}\log\gamma+\sum_{i=1}^{3}\log\left(\frac{\lambda_i}{\gamma}+1\right)\,. \tag{3.25}$$

Computing the derivative of $\log f_3(\gamma)$, $\gamma\in]0\,,+\infty[$, and equating it to zero yields

$$\frac{3}{2}-\sum_{i=1}^{3}\frac{\lambda_i}{\lambda_i+\gamma}=0\,, \tag{3.26}$$

which is tantamount to solving

$$\gamma^3+\frac{1}{3}(\lambda_1+\lambda_2+\lambda_3)\gamma^2-\frac{1}{3}(\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3)\gamma-\lambda_1\lambda_2\lambda_3=0\,. \tag{3.27}$$

This is a third order equation with real coefficients. Descartes' rule of signs implies that it shares one positive real root which necessarily coincides with the optimal point of (3.23). Additionally, Tartaglia formula can be exploited to get the analytic expression of the optimizer $\gamma_{opt,3}$. Precisely, denoting by $a=\frac{1}{3}(\lambda_1+\lambda_2+\lambda_3)$, $b=-\frac{1}{3}(\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3)$, $c=-\lambda_1\lambda_2\lambda_3$, $p=-a^2/3+b$, $q=2a^3/27-ab/3+c$, the optimal point is the real positive number chosen among the three roots of (3.27) given by

$$-\frac{a}{3}+\sqrt[3]{-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}$$

where $\sqrt[3]{(\cdot)}$ and $\sqrt{(\cdot)}$ are complex roots[2].

[2]Given a complex number $y=\rho\exp(j\theta)$ the n-th roots are $u_k=\rho^{\frac{1}{n}}\left[\cos\left(\frac{\theta}{n}+\frac{2\pi k}{n}\right)+j\cos\left(\frac{\theta}{n}+\frac{2\pi k}{n}\right)\right]$, $k=0,1,\ldots,n-1$.

Summarizing, the GLRT can be computed as

$$\frac{\left[\gamma_{opt,3}^{3}\left(\frac{\lambda_1}{\gamma_{opt,3}}+1\right)^{2}\left(\frac{\lambda_2}{\gamma_{opt,3}}+1\right)^{2}\left(\frac{\lambda_3}{\gamma_{opt,3}}+1\right)^{2}\right]}{\lambda_1\lambda_2\lambda_3} \underset{H_0}{\overset{H_1}{\gtrless}} T\,. \tag{3.28}$$

An equivalent expression in terms of the maximal invariant can be also obtained. From (3.27), it is not difficult to prove that

$$\gamma_{opt,3} = \lambda_1 h\left(\frac{\lambda_2}{\lambda_1}, \frac{\lambda_3}{\lambda_1}\right), \tag{3.29}$$

with $h(\cdot,\cdot)$ a suitable bi-dimensional real function of two real variables. Hence, the GLRT statistic as function of the maximal invariant can be obtained substituting (3.29) in (3.28).

3.3.3 Additional Suboptimum Invariant Detectors

This section presents additional sub-optimum invariant detectors with reference to the fully polarimetric processing (i.e. $N = 3$). They exploit suitable combinations of the maximal invariant components and, based on extensive numerical analysis, are seen to achieve satisfactory detection performances. Two chosen combination rules are the standard arithmetic mean and geometric mean which respectively lead to the tests

$$\frac{\lambda_1}{\lambda_2}+\frac{\lambda_1}{\lambda_3} \underset{H_0}{\overset{H_1}{\gtrless}} T_a, \tag{3.30}$$

$$\frac{\lambda_1}{\lambda_2}\frac{\lambda_1}{\lambda_3} \underset{H_0}{\overset{H_1}{\gtrless}} T_b, \tag{3.31}$$

where T_a and T_b are the detection thresholds set to ensure a specified level of P_{fa}.

Moreover, based on the observations that unbalances among the eigenvalues of $\boldsymbol{\Sigma}_Y^{-\frac{1}{2}}\boldsymbol{\Sigma}_X\boldsymbol{\Sigma}_Y^{-\frac{1}{2}}$ are representative of changes between the reference and test images; and an index to quantify eigenvalues' unbalances is the deviation of the arithmetic mean from the geometric mean[3],

[3] If $\boldsymbol{A}$ is a positive definite matrix $N \times N$ matrix, the following inequality holds true $[\det(\boldsymbol{A})]^{\frac{1}{N}} \leq \frac{1}{N}\mathrm{tr}\,(\boldsymbol{A})$. Equality holds if and only if $\boldsymbol{A} = \boldsymbol{I}$ [28, p. 477].

the following decision rule is also considered

$$\frac{\mathrm{tr}\left(\boldsymbol{S}_Y^{-\frac{1}{2}}\boldsymbol{S}_X\boldsymbol{S}_Y^{-\frac{1}{2}}\right)}{\det^{\frac{1}{3}}\left(\boldsymbol{S}_Y^{-\frac{1}{2}}\boldsymbol{S}_X\boldsymbol{S}_Y^{-\frac{1}{2}}\right)} = \left(\frac{\lambda_2}{\lambda_1}\frac{\lambda_3}{\lambda_1}\right)^{-\frac{1}{3}}\left(1+\frac{\lambda_2}{\lambda_1}+\frac{\lambda_3}{\lambda_1}\right) \underset{H_0}{\overset{H_1}{\gtrless}} T_c, \quad (3.32)$$

where T_c is the detection threshold.

In the next section the performance of the previously introduced invariant (and hence CFAR) decision rules is compared with the GLRT (3.28) both on simulated and on real data.

3.4 Performance Analysis

This section presents the performance analysis of the proposed scale-invariant detectors for both simulated and real data. In particular, P_{fa} and P_d are obtained through Monte Carlo simulations[4]. Then real data is used to demonstrate the capability of the new algorithms to operate in real challenging environments.

3.4.1 Performance Analysis on Simulated Data

This section presents the performance analysis via computer simulated data of the detectors introduced in Section IV. Three different studies are performed to assess the properties of the proposed receivers.

The first is conducted in terms of P_d for a given P_{fa} level, assuming zero-mean complex circular multivariate Gaussian observations with equal (but not proportional) covariance matrices under H_0 (H_1). By doing so, all the information characterizing the set of polarimetric SAR images is contained in the covariance matrix. This is equivalent to assigning a) diagonal elements: power information, b) off-diagonal elements: correlation information. Different number of polarizations and size of the square inspection window are considered. Monte Carlo simulation is used to set the detection thresholds assuming $100/P_{fa}$ independent runs and considering a $P_{fa} = 10^{-4}$. The value of P_d is estimated using 5000 independent trials.

The first study refers to $N = 2$, in this case the maximal invariant,

[4]The only exception where an analytic performance derivation can be carried on is the GLRT for $N = 2$ (3.21). See Appendix D for details.

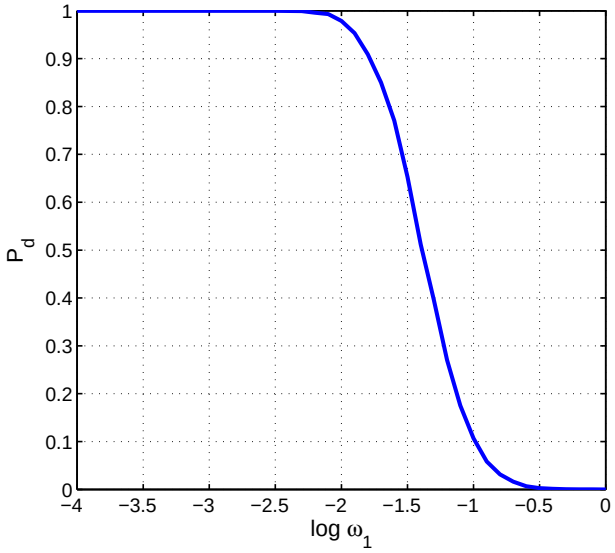
which completely governs the performance of any invariant detector, is one-dimensional. This means that the P_d plots versus the induced maximal invariant $\omega_1 = \delta_2/\delta_1 \leq 1$ completely characterize the performance of detector (3.21). Otherwise stated, the reported performances holds for any pair of covariances $(\boldsymbol{\Sigma}_X\,;\,\boldsymbol{\Sigma}_Y)$ sharing the same induced maximal invariant. This is a consequence of the fact that the performance of invariant detectors depends on the pair $(\boldsymbol{\Sigma}_X\,;\,\boldsymbol{\Sigma}_Y)$ only through the induced maximal invariant.

In Figure 3.1-a and Figure 3.1-b the values of P_d for $W = 3$ and $W = 5$ are drawn versus ω_1. For $W = 3$ the receiver (3.21) provides a $P_d \geq 0.9$ for $\omega_1 \notin \left[10^{-1.8}\,,\,1\right]$. The case $W = 5$ makes use of a greater number of homogeneous data vectors in the Grammians $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ whose scaled versions ($1/K$ scale factor) are unbiased and consistent estimates of the covariance matrices $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$. For this situation a $P_d \geq 0.9$ is achieved for $\omega_1 \notin \left[10^{-1.02}\,,\,1\right]$. A similar analysis is performed for $N = 3$ and $W = 3$, shown in Figures 3.2 and 3.3. In this case the contours showing P_d are functionally dependent on two variables (namely the two components of the induced maximal invariant $\boldsymbol{\omega}$): $\omega_1 = \delta_2/\delta_1 \leq 1$, and $\omega_2 = \delta_3/\delta_1 \leq 1$. In Figure 3.2-a the contour plot for detector (3.28) is shown; for this scenario $P_d \geq 0.9$ is guaranteed if $\omega_i \notin [10^{-2.11}, 1]$, $i = 1, 2$. The results for detector (3.30) are shown in Figure 3.2-b, in this case an acceptable P_d is achieved for $\omega_i \notin [10^{-2.14}, 1]$, $i = 1, 2$. For the decision rules (3.31) and (3.32), whose results are displayed in Figure 3.3-a and Figure 3.3-b, a $P_d \geq 0.9$ is guaranteed for $\omega_i \notin [10^{-2.83}, 1]$, and $\omega_i \notin [10^{-2.46}, 1]$, $i = 1, 2$, respectively. The contour plots for $N = 3$ and $W = 5$ are shown in Figures 3.4 and 3.5; the main difference with the case $W = 3$ is the increase in the detection performance, as observed for the $N = 2$ analysis. In particular the region in the (ω_1, ω_2) space where P_d is greater than 0.9 grows, for example for the decision rule (3.28) the region is now described by the values $\omega_i \notin [10^{-1.1}, 1]$, $i = 1, 2$.

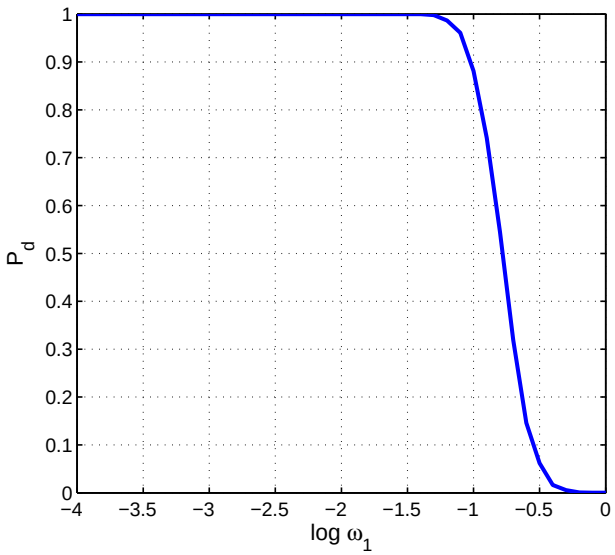
The second analysis is conducted in terms of ROC curves (namely P_d versus P_{fa}) for a fixed value of the induced maximal invariant $\boldsymbol{\omega} = [0.04, 0.03]^T$, corresponding to the pair of covariance matrices:

$$\boldsymbol{C}_1 = \begin{pmatrix} 16 & 0 & 0.7 \\ 0 & 0.2 & 0 \\ 0.7 & 0 & 1 \end{pmatrix} \qquad \boldsymbol{C}_2 = 2 \begin{pmatrix} 4 & 0 & 0.1 \\ 0 & 1.5 & 0 \\ 0.1 & 0 & 6 \end{pmatrix}.$$

Of course, any other covariance pair with the same value of $\boldsymbol{\omega}$ leads to the same ROCs, while a different value of $\boldsymbol{\omega}$ leads to different ROCs.

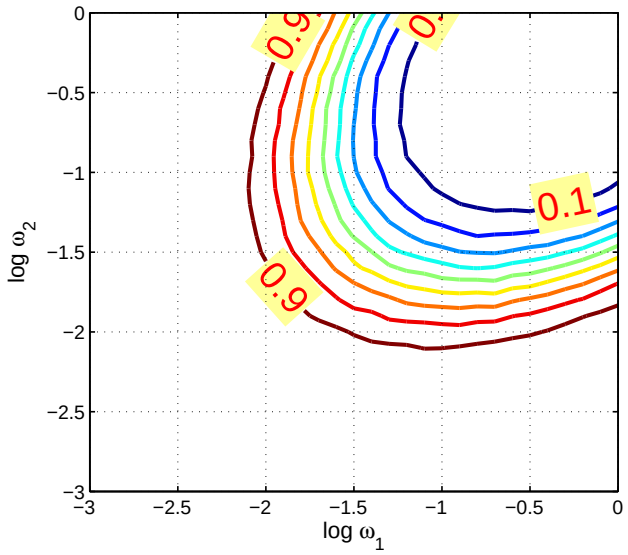


(a)

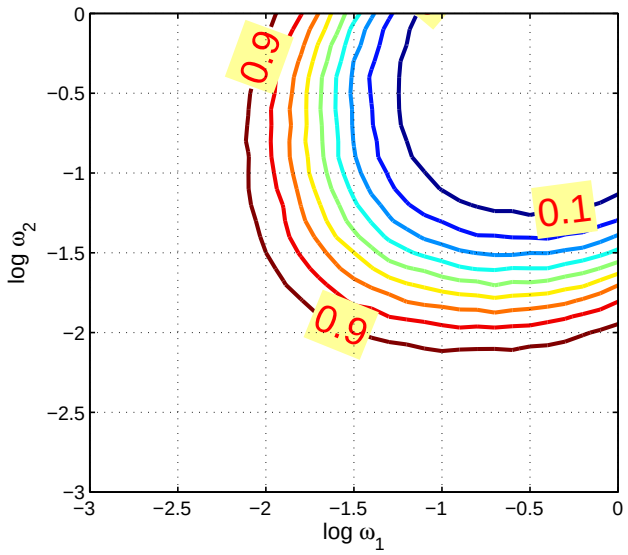


(b)

Figure 3.1: a) P_d versus $\log \omega_1$, $N = 2$ and $W = 3$; b) P_d versus $\log \omega_1$, $N = 2$ and $W = 5$.

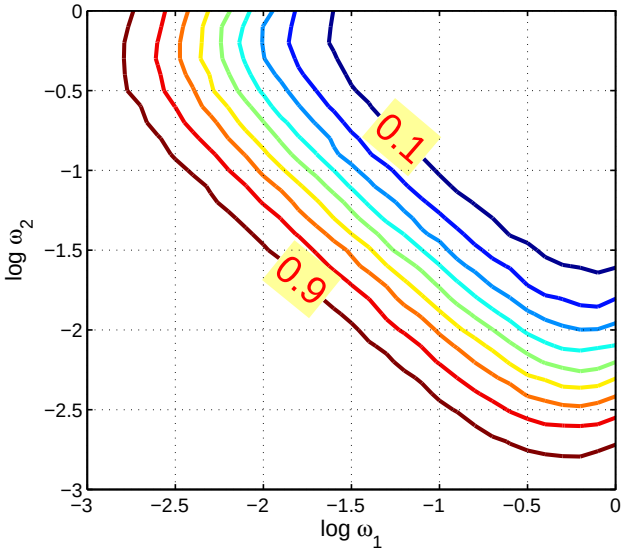


(a) Detector (3.28)

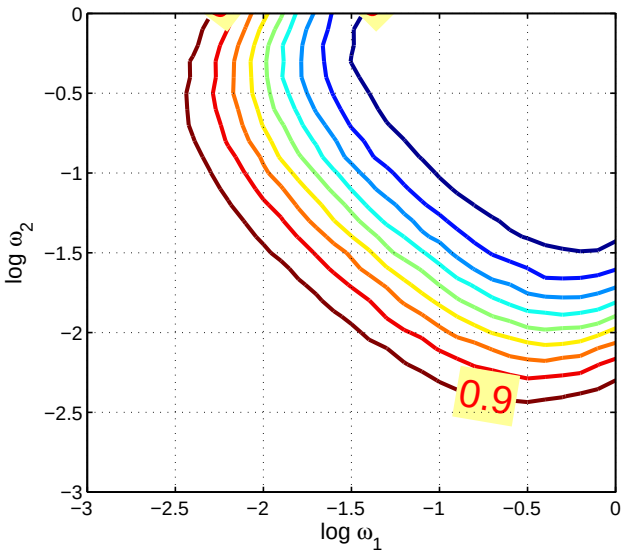


(b) Detector (3.30)

Figure 3.2: P_d contours versus $\log \omega_1$ and $\log \omega_2$ for $N = 3$ and $W = 3$.

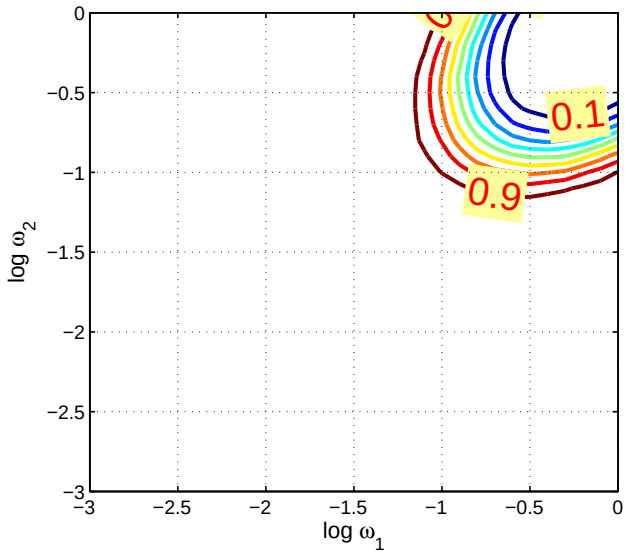


(a) Detector (3.31)

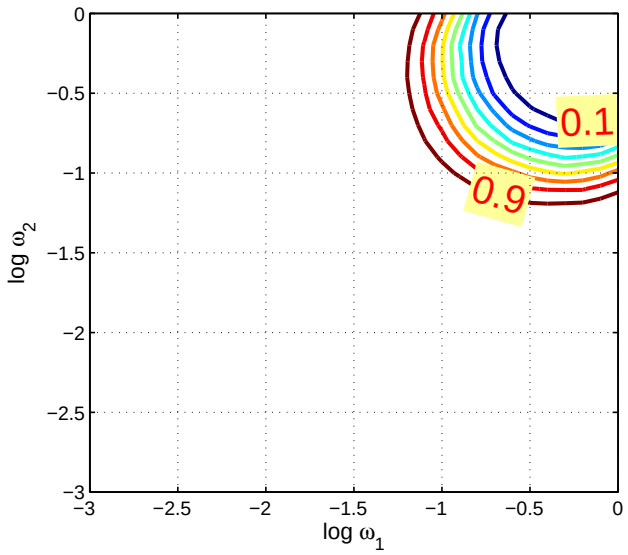


(b) Detector (3.32)

Figure 3.3: P_d contours versus $\log \omega_1$ and $\log \omega_2$ for $N = 3$ and $W = 3$.

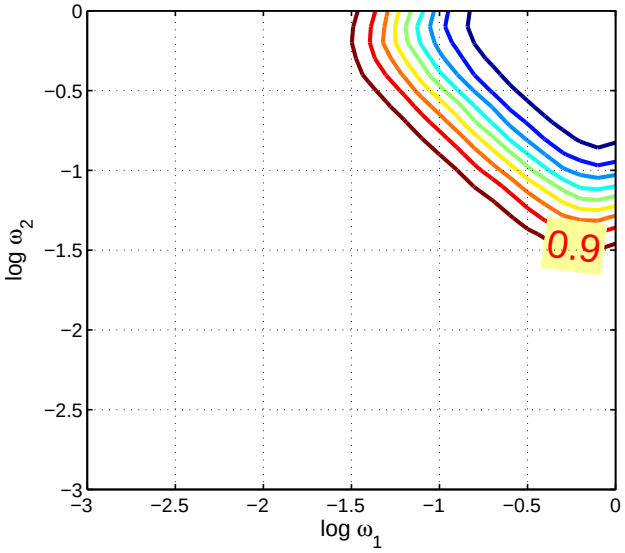


(a) Detector (3.28)

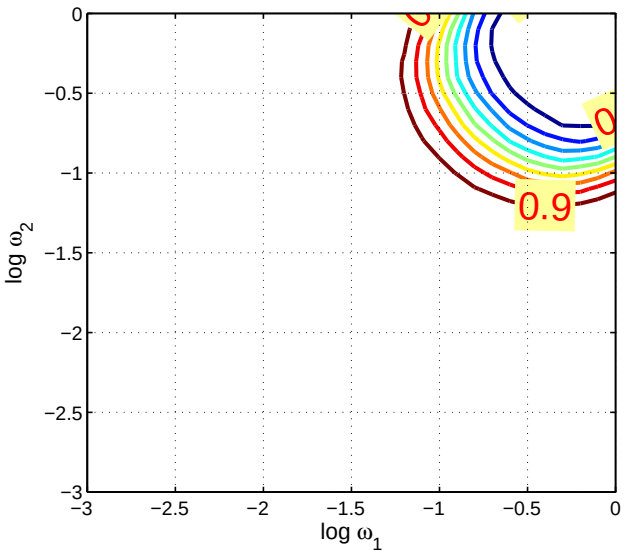


(b) Detector (3.30)

Figure 3.4: P_d contours versus $\log \omega_1$ and $\log \omega_2$ for $N = 3$ and $W = 5$.



(a) Detector (3.31)



(b) Detector (3.32)

Figure 3.5: P_d contours versus $\log \omega_1$ and $\log \omega_2$ for $N = 3$ and $W = 5$.

For comparison purposes, the ROC of the detector (1.20) and of that in [9] are also reported. Specifically, the decision rule proposed in [9] is an adaptive implementation of the LRT, whose expression in terms of the eigenvalues $\lambda_1, \ldots, \lambda_N$ is

$$\sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \ln \frac{1}{\lambda_i} \right) \underset{H_0}{\overset{H_1}{\gtrless}} T_H \,, \tag{3.33}$$

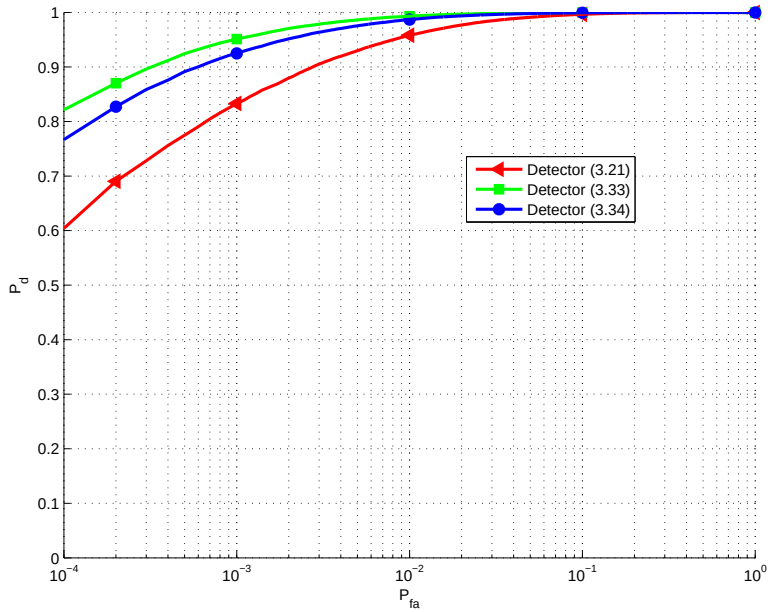
whereas the GLRT (1.20) is

$$\prod_{i=1}^{N} \frac{(1+\lambda_i)^2}{\lambda_i} \underset{H_0}{\overset{H_1}{\gtrless}} T_g \,. \tag{3.34}$$

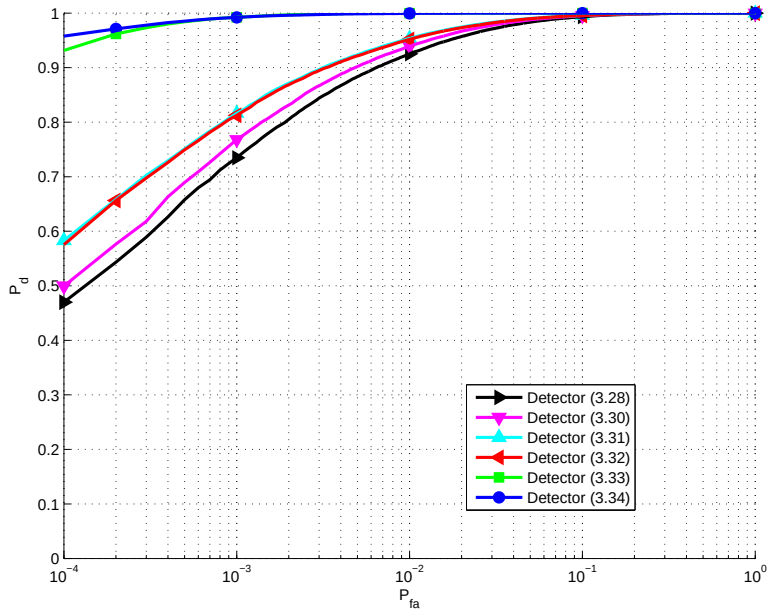
In order to set the detection thresholds, Monte Carlo simulations were used assuming $100/P_{fa}$ independent runs. Additionally, 100000 independent trials were exploited to estimate the P_d.

Figure 3.6 shows the ROCs of the considered receivers for both two- and three-polarimetric channels and $W = 3$. In particular Figure 3.6-a shows the two-polarization case while Figure 3.6-b refers to the $N = 3$ case. In all the analyzed situations, the ROC highlights that the scale-invariant detectors are outperformed in terms of P_d by the two receivers proposed in Chapter 1 and in [9] if $W = 3$. However, as soon as $W = 5$ all the receivers provide the unit P_d value for $P_{fa} \in [10^{-4}, 1]$. This behaviour is actually expected as, at the design stage, we are requiring "more invariance" than the detectors in Chapter 1 and in [9]. In fact, the new decision rules exhibit an additional scale invariance attempting to robustify the CFAR property with respect to possible scale mismatches between the reference and test images. In other words, as it will also be clearer at the end of this sub-section, with the new approach we are trading off detection performance with an improved CFAR behavior.

The final analysis assesses the benefits of the proposed approach in terms of CFAR property. The study is conducted considering a nominal $P_{fa} = 10^{-4}$ in the presence of observations with equal covariance matrices (i.e. assuming $C_1 = C_2$). To estimate the actual P_{fa} the covariance matrix C_1 is the same used for the ROC analysis while C_2 is selected as $C_2 = \alpha C_1$ with $\alpha \in [0.5, 2]$. The nominal threshold for a $P_{fa} = 10^{-4}$ is then used to estimate the actual P_{fa} for each detector.



(a) N=2, W=3



(b) N=3, W=3

Figure 3.6: P_d versus P_{fa}.

In Figures 3.7 and 3.8, respectively, P_{fa} versus α for $N = 2$ and $N = 3$ with inspection window sizes $W = 3$ and $W = 5$, is shown. Also for this analysis the curve for the detector (3.34) and for the detector (3.33) proposed in [9] are represented. In all the considered situations the invariant detectors show a stable actual probability of false alarm at 10^{-4} (all their P_{fa} curves overlap with the horizontal line at 10^{-4}). The detectors (3.34) and (3.33) exhibit poor capability in handling scale variations between the reference and test images, providing an actual P_{fa} different from the nominal value when a scale variation is present ($\alpha \neq 1$). For example, in the case $N = 3$ and $W = 5$, the detector in (3.34) exhibits an actual P_{fa} of 0.139 for $\alpha = 2$ while the proposed invariant rules are still able to perform with an actual P_{fa} of 10^{-4} as predicted by the developed theory.

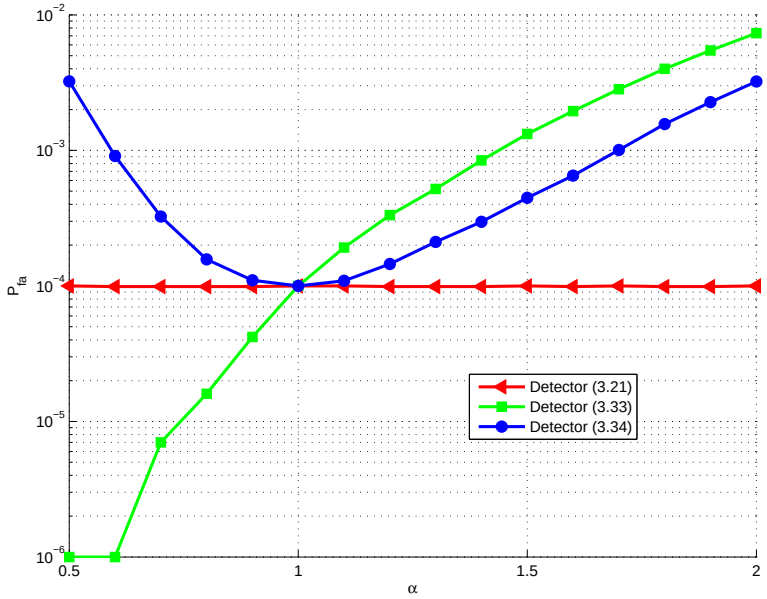
3.4.2 Performance Analysis on Real Data

In the following a CFAR analysis based on the previously described real data is conducted. To this end, the thresholds are set to ensure $P_{fa} = 10^{-4}$ in the complement of the extended ground truth area, namely, in the region where no changes occur (there are no true positives). This means that, for each detector, after computing the decision statistics (for each pixel belonging to the complement of the extended ground truth), the threshold has been selected in order that

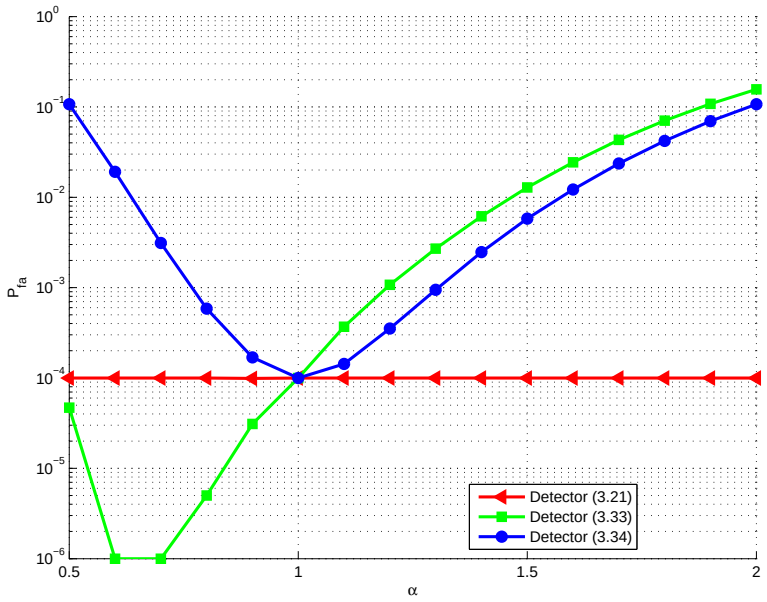
$$10^{-4} \times \text{total number of available statistics (trials)},$$

are greater than the threshold. This ensures that all the comparisons refer to the same nominal P_{fa} level, namely the number of threshold crossings in the complement of the extended ground truth is exactly the same for all the analysed detectors. Then exploiting the previously computed thresholds the actual P_{fa} is estimated applying a scaling $\sqrt{\alpha}$ to the test image for $\alpha \in [0.5, 2]$.

In Figures 3.9 and 3.10, respectively, the P_{fa} versus α for $N = 2$ and $N = 3$ with inspection window sizes of $W = 3$ and $W = 5$, is shown. The behaviours of the detector (3.34) and of the detector (3.33) are also displayed.

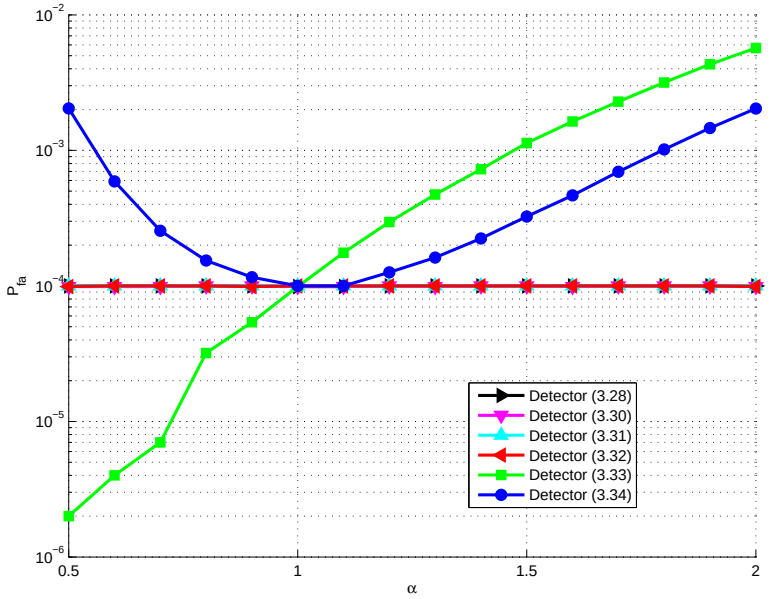


(a) $W = 3$.

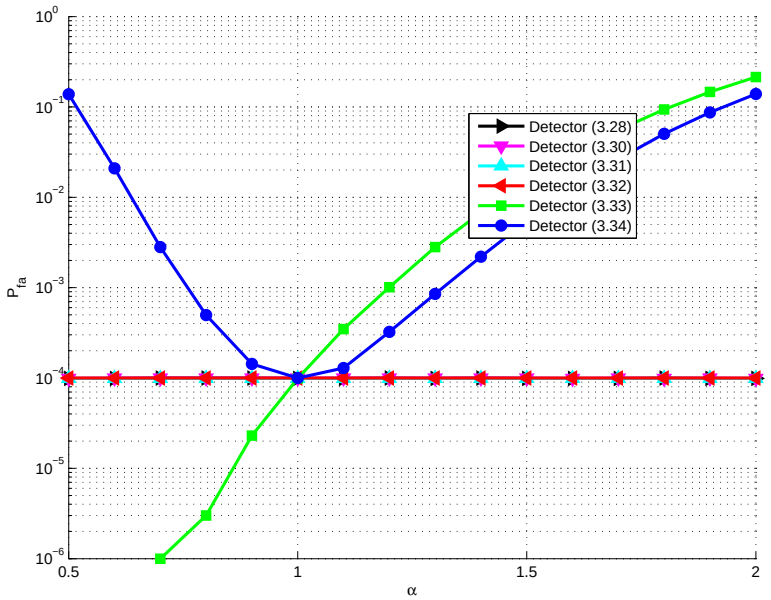


(b) $W = 5$.

Figure 3.7: P_{fa} versus α considering $N = 2$.

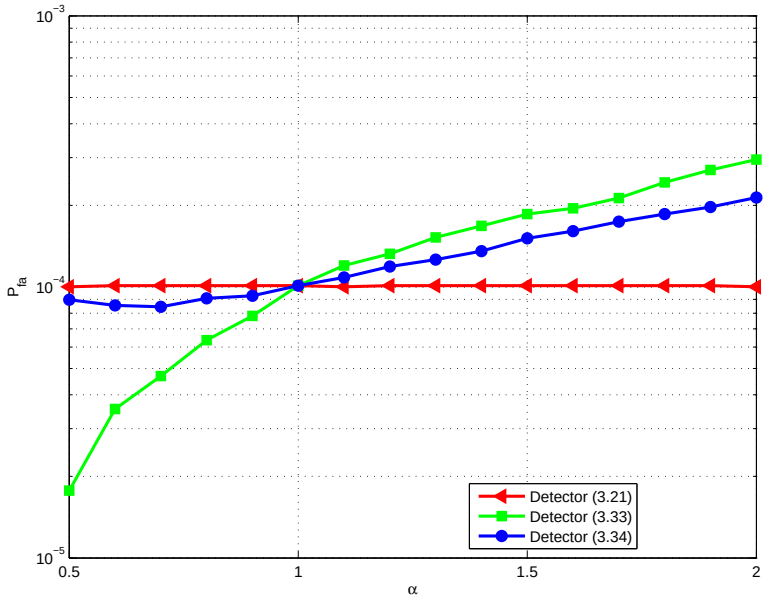


(a) $W = 3$.

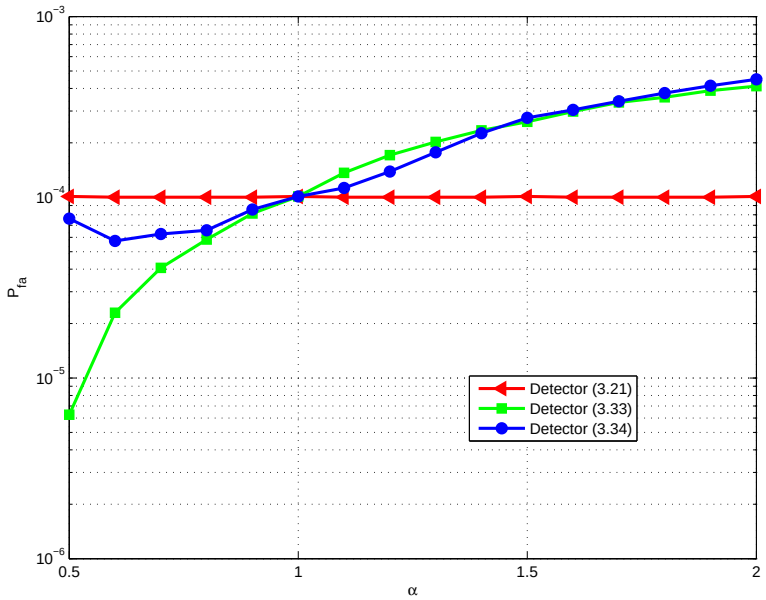


(b) $W = 5$.

Figure 3.8: P_{fa} versus α considering $N = 3$.

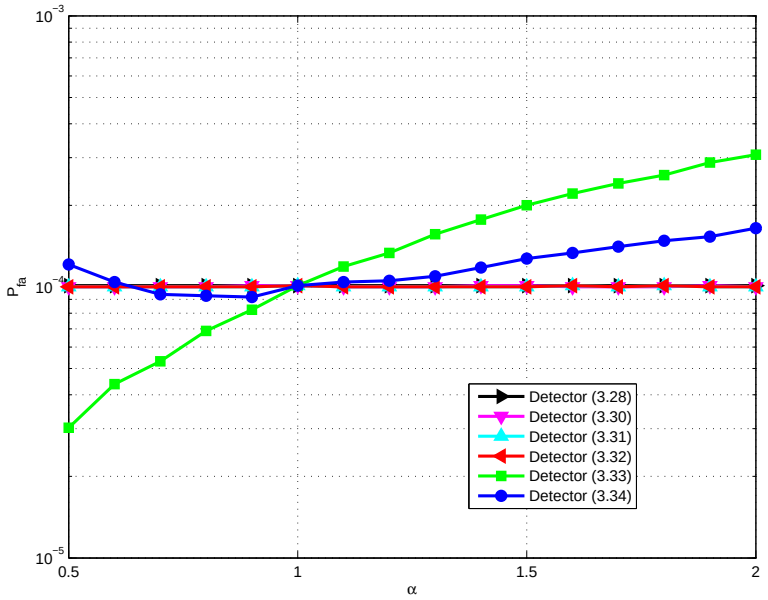


(a) $W = 3$.

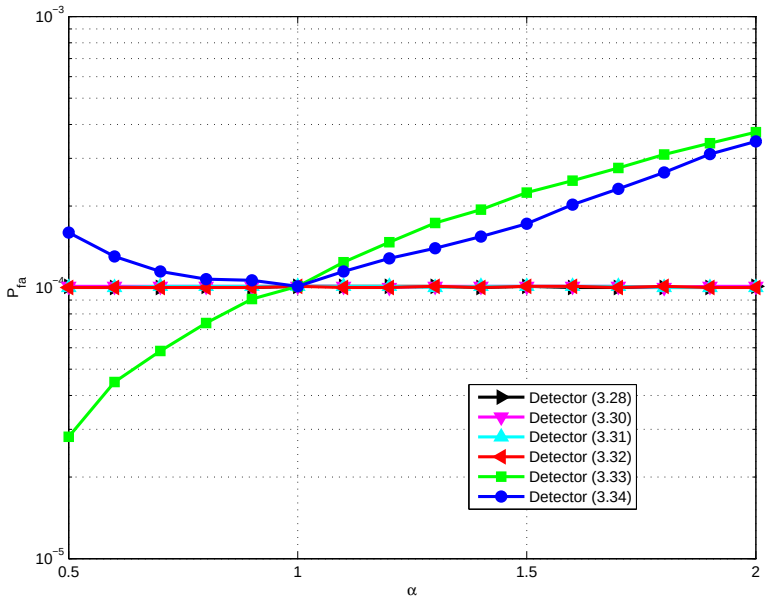


(b) $W = 5$.

Figure 3.9: P_{fa} versus α considering $N = 2$.



(a) $W = 3$.



(b) $W = 5$.

Figure 3.10: P_{fa} versus α considering $N = 3$.

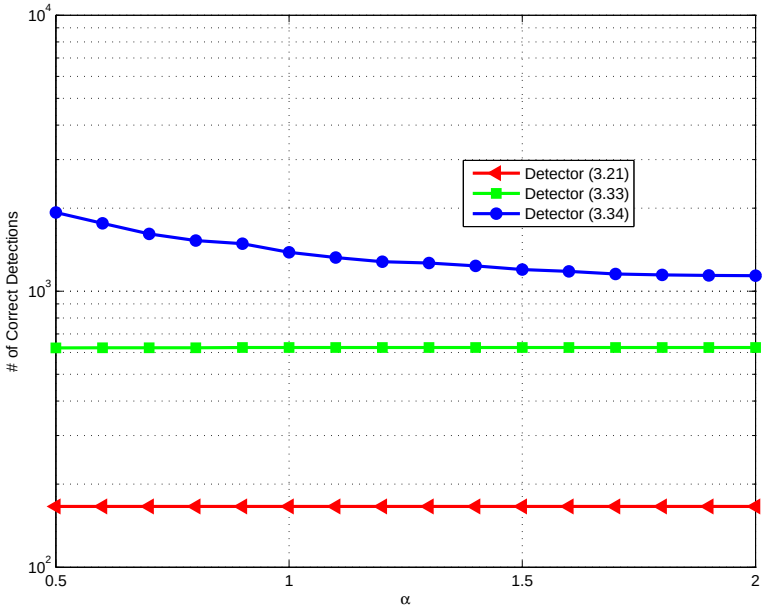
Confirming the simulated analysis, in all the situations, the invariant detectors show to provide a stable probability of false alarm. The detectors (3.34) and (3.33), show poor ability to keep a constant false alarm rate in the presence of scale variations between the reference and test images. Specifically, they exhibit an actual P_{fa} different from the nominal value when a scale variation is present ($\alpha \neq 1$).

For the same scenarios the number of correct detections in the extended ground truth has been evaluated. The results are shown in Figures 3.11 and 3.12; as expected, the scale-invariant detectors provide constant performance for different values of α, but in this case they are outperformed by the detectors (3.34) and (3.33).

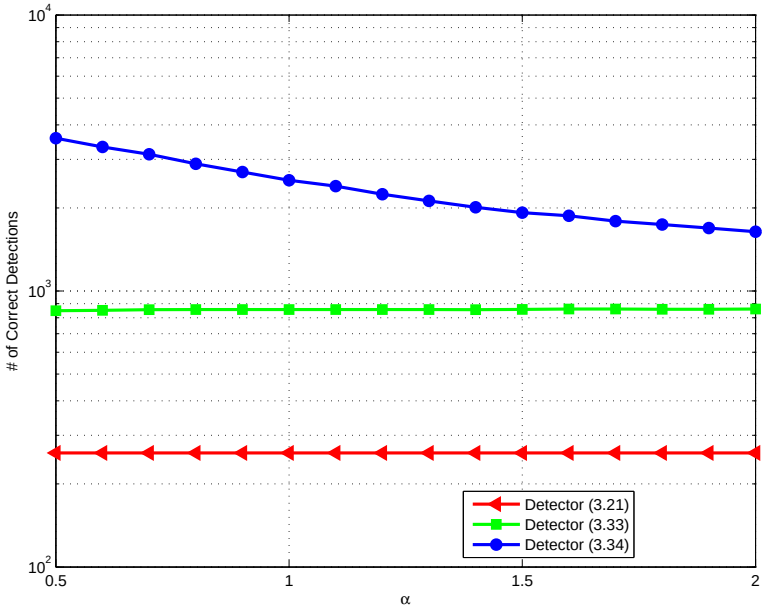
The final analysis investigates the effects of an aggregation strategy after single pixel detection in order to eliminate isolated false alarms and confirm true detections, which are to the typical car size and system resolution, appear quite often. More precisely, a window of size 5×5 slides along the horizontal and the vertical dimensions of the detection maps (which, as already highlighted are binary images: "0" no detection, "1" detection). Then, for a given pixel localized at the center of the moving window and labelled with "1", a detection is associated if the number of "1" in the window is greater that a certain integer "fill parameter" (denoted by F and complying with $1 \leq F \leq 25$).

These kind of logics look like n-of-m aggregation techniques [34, Ch. 3] and can be interpreted as a post-detection binary integration within the reference window. As to the pixels on the image edge (namely those laying on the first/last two rows or columns), no aggregation is performed because they never fall at the center of the moving window. In other words, it is simply confirmed the "0" or "1" value in the original detection map.

The case of three-polarimetric channels, with $W = 5$ and $\alpha = 1$ is considered and the effect of the fill parameter is studied in Figure 3.13 for detectors (3.28), (3.30), (3.31) and (3.32). This analysis highlights that all the proposed receivers are sensitive to the fill parameter value. Otherwise stated to get a lower number of false alarms through the aggregation procedure, it is necessary to accept some detection loss.

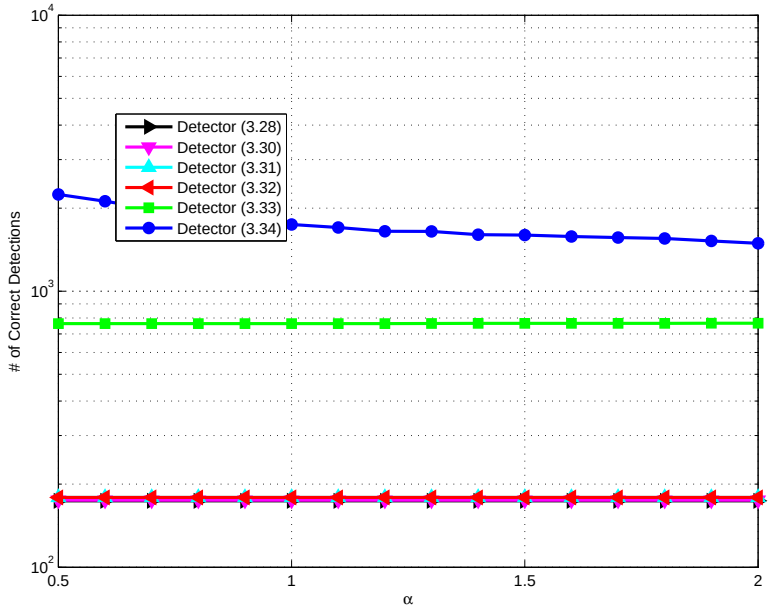


(a) $W = 3$.

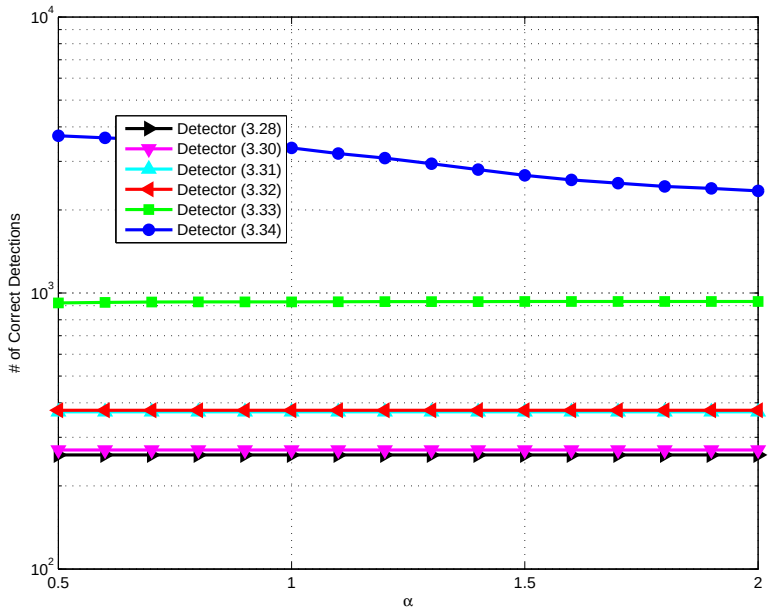


(b) $W = 5$.

Figure 3.11: # of Correct Detections versus α considering $N = 2$.

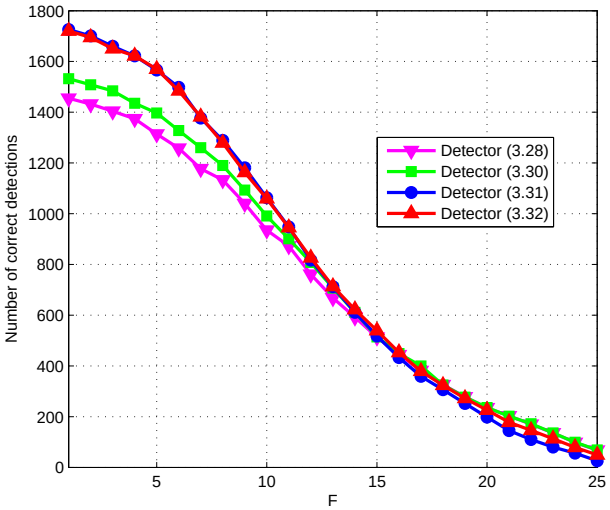


(a) $W = 3$.

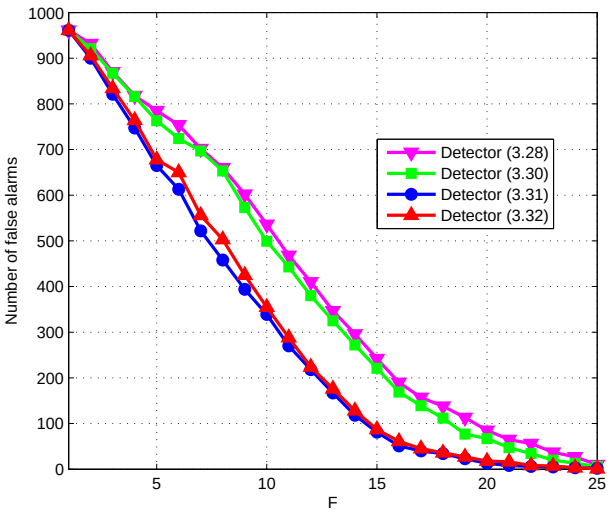


(b) $W = 5$.

Figure 3.12: # of Correct Detections versus α considering $N = 3$.



(a) Correct Detections



(b) False Alarms

Figure 3.13: Number of correct detections in the extended ground truth (a) and number of false alarms in the complement of the extended ground truth (b) versus F.

3.5 Conclusions

In this chapter multi-polarization scale-invariant change detection from SAR images has been considered. The problem has been formulated as a binary hypothesis test and the principle of invariance has been applied to synthesize decision rules. This framework allows to both enhance the robustness of the detectors with respect to intensity miscalibration effects and to force the CFAR feature at the design stage. A maximal invariant statistic, which fully characterizes the class of invariant tests, as well as the induced maximal invariant were determined with reference to the processing of two- or three-polarimetric channels. Moreover, the GLRT is computed; interestingly for two-polarimetric channels the test is equivalent to comparing the condition number of a data-dependent matrix with a suitable detection threshold. Further decision rules have been also introduced as combinations of the maximal invariant components.

At the analysis stage, the proposed framework has been assessed both on simulated data and on real high resolution SAR images. The conducted study has shown the capability of the novel decision rules to provide the CFAR property even in the presence of power mismatches among the different acquisitions. The tradeoff between robustness and detection performance has also been discussed. Future work will perform an analysis of the new detectors as well as of possible polarimetric extensions of other decision rules such as that in [36], in the presence of other real datasets collected under different environmental and operating conditions.

Appendix A

In this appendix the MPI detector is derived with reference to the case of $N = 3$ polarimetric channels. To this end, it is necessary to distinguish among three cases.

- **Case 1**: $\delta_1 \neq \delta_2 \neq \delta_3$. Using [30, eq. (51)], ${}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda})$ can be recast as a rational function

$$ {}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda}) = \frac{c_3 \det\left[\left(1+\frac{\lambda_j}{\delta_i}\right)^{2-2K}\right]_{i,j=1,\ldots,3}}{\prod_{j=1}^{3}\prod_{i<j}\left(-\frac{1}{\delta_i}+\frac{1}{\delta_j}\right)(\lambda_i-\lambda_j)} \tag{A.1} $$

  with c_3 a constant with respect to λ_i and δ_i. Thus the MPI detector becomes

$$ \frac{\det\left[\left(1+\frac{\lambda_j}{\delta_i}\right)^{2-2K}\right]_{i,j=1,\ldots,3}}{\prod_{j=1}^{3}(1+\lambda_j)^{-2K}\prod_{i<j}\left(-\frac{1}{\delta_i}+\frac{1}{\delta_j}\right)(\lambda_i-\lambda_j)} \underset{H_0}{\overset{H_1}{\gtrless}} T_{A,3}\,. \tag{A.2} $$

- **Case 2**: $\delta_1 = \delta_2 \neq \delta_3$ or $\delta_1 \neq \delta_2 = \delta_3$ or $\delta_1 = \delta_3 \neq \delta_2$. Let us focus on $\delta_1 = \delta_2 \neq \delta_3$ (the other two cases are equivalent to the considered one). Denoting by $\delta = \delta_1 = \delta_2$ and exploiting [37, eq. (106)] the following equality chain holds true

$$
{}_1\tilde{F}_0(2K;-\boldsymbol{\Delta}^{-1},\boldsymbol{\Lambda}) = \frac{c_3}{\prod_{j=1}^{3}\prod_{i<j}(\lambda_i-\lambda_j)}\times
$$

$$
\times \lim_{\substack{x_1\to -\frac{1}{\delta}\\ x_2\to -\frac{1}{\delta}}} \frac{\det\begin{bmatrix} (1-\lambda_1 x_1)^{2-2K} & (1-\lambda_1 x_2)^{2-2K} & \left(1+\frac{\lambda_1}{\delta_3}\right)^{2-2K} \\ (1-\lambda_2 x_1)^{2-2K} & (1-\lambda_2 x_2)^{2-2K} & \left(1+\frac{\lambda_2}{\delta_3}\right)^{2-2K} \\ (1-\lambda_3 x_1)^{2-2K} & (1-\lambda_3 x_2)^{2-2K} & \left(1+\frac{\lambda_3}{\delta_3}\right)^{2-2K} \end{bmatrix}}{(x_2-x_1)\left(-x_1-\frac{1}{\delta_3}\right)\left(-x_2-\frac{1}{\delta_3}\right)} =
$$

$$
c_3 \frac{\det\begin{bmatrix} = \left(1+\frac{\lambda_1}{\delta}\right)^{2-2K} & -\lambda_1(2-2K)\left(1+\frac{\lambda_1}{\delta}\right)^{1-2K} & \left(1+\frac{\lambda_1}{\delta_3}\right)^{2-2K} \\ \left(1+\frac{\lambda_2}{\delta}\right)^{2-2K} & -\lambda_2(2-2K)\left(1+\frac{\lambda_2}{\delta}\right)^{1-2K} & \left(1+\frac{\lambda_2}{\delta_3}\right)^{2-2K} \\ \left(1+\frac{\lambda_3}{\delta}\right)^{2-2K} & -\lambda_3(2-2K)\left(1+\frac{\lambda_3}{\delta}\right)^{1-2K} & \left(1+\frac{\lambda_3}{\delta_3}\right)^{2-2K} \end{bmatrix}}{\left(\frac{1}{\delta}-\frac{1}{\delta_3}\right)^2 \prod_{j=1}^{3}\prod_{i<j}(\lambda_i-\lambda_j)}. \tag{A.3}
$$

As a consequence the MPI test becomes

$$
\frac{\det\begin{bmatrix} \left(1+\frac{\lambda_1}{\delta}\right)^{2-2K} & -\lambda_1(2-2K)\left(1+\frac{\lambda_1}{\delta}\right)^{1-2K} & \left(1+\frac{\lambda_1}{\delta_3}\right)^{2-2K} \\ \left(1+\frac{\lambda_2}{\delta}\right)^{2-2K} & -\lambda_2(2-2K)\left(1+\frac{\lambda_2}{\delta}\right)^{1-2K} & \left(1+\frac{\lambda_2}{\delta_3}\right)^{2-2K} \\ \left(1+\frac{\lambda_3}{\delta}\right)^{2-2K} & -\lambda_3(2-2K)\left(1+\frac{\lambda_3}{\delta}\right)^{1-2K} & \left(1+\frac{\lambda_3}{\delta_3}\right)^{2-2K} \end{bmatrix}}{\prod_{j=1}^{3}(1+\lambda_j)^{-2K}\prod_{i<j}(\lambda_i-\lambda_j)} \underset{H_0}{\overset{H_1}{\gtrless}} T_{B,3}\,. \tag{A.4}
$$

- **Case 3**: $\delta_1 = \delta_2 = \delta_3$. Using the splitting formula [26, eq. (92)], yields,

$$ {}_1\tilde{F}_0(2K; -\boldsymbol{\Delta}^{-1}, \boldsymbol{\Lambda}) = \det\left(\boldsymbol{I} + \frac{\boldsymbol{\Lambda}}{\delta}\right)^{-2K}, \quad \delta_1 = \delta_2 = \delta_3 = \delta\,. \tag{A.5} $$

Hence, after standard equivalences, the MPI test (1.12) becomes

$$ \prod_{j=1}^{3}\left(\frac{1+\lambda_j}{\delta+\lambda_j}\right) \underset{H_0}{\overset{H_1}{\gtrless}} T_{C,3}\,. \tag{A.6} $$

Appendix B

To prove the CFAR property it is sufficient to observe that, under H_0,

- $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ are identically distributed complex Wishart random matrices with scale parameter K and matrix parameter $\boldsymbol{\Sigma}_X$ [26];
- $\hat{\sigma}^2_{X,1}$ and $\hat{\sigma}^2_{Y,1}$ admit the following stochastic representation:

$$\hat{\sigma}^2_{X,1} = \sigma^2 \chi_1, \quad \hat{\sigma}^2_{Y,1} = \sigma^2 \chi_2 \tag{B.1}$$

  with χ_1 and χ_2 two complex Chi-square random variables with K degrees of freedom.

Moreover $\boldsymbol{S}_X$, $\boldsymbol{S}_Y$, χ_1, and χ_2 are statistically independent.

It follows that the left hand side of (2.10) can be rewritten as

$$\frac{\det^2 \left[\boldsymbol{\Sigma}_X^{\frac{1}{2}} \left(\tilde{\boldsymbol{S}}_{X,1} + \tilde{\boldsymbol{S}}_{Y,1} \right) \boldsymbol{\Sigma}_X^{\frac{1}{2}} \right]}{\det \left(\boldsymbol{\Sigma}_X^{\frac{1}{2}} \tilde{\boldsymbol{S}}_{X,1} \boldsymbol{\Sigma}_X^{\frac{1}{2}} \right) \det \left(\boldsymbol{\Sigma}_X^{\frac{1}{2}} \tilde{\boldsymbol{S}}_{Y,1} \boldsymbol{\Sigma}_X^{\frac{1}{2}} \right)} \frac{(\chi_1 + \chi_2)^2}{\chi_1 \chi_2}, \tag{B.2}$$

or equivalently as

$$\frac{\det^2 \left(\tilde{\boldsymbol{S}}_{X,1} + \tilde{\boldsymbol{S}}_{Y,1} \right)}{\det \left(\tilde{\boldsymbol{S}}_{X,1} \right) \det \left(\tilde{\boldsymbol{S}}_{Y,1} \right)} \frac{(\chi_1 + \chi_2)^2}{\chi_1 \chi_2}, \tag{B.3}$$

with $\tilde{\boldsymbol{S}}_{X,1}$ and $\tilde{\boldsymbol{S}}_{Y,1}$ two independent and identically distributed complex Wishart random matrices with scale parameter K and identity matrix parameter.

The stochastic representation in (B.3) highlights that the decision statistic in (2.10) can be expressed in terms of random variables whose marginal pdf (and, hence, also their joint pdf due to the statistical independence) is functionally independent of $\mathbf{\Sigma}_X$. This implies that the CFAR property is achieved.

Appendix C

In this appendix we show that the set of transformations G is a group. To this end we define the *group multiplication* [16, p. 569] which with any two elements of G

$$g_1 = [\boldsymbol{B}_1, a_1 \,:\, \boldsymbol{B}_1 \in \mathcal{GL}(N),\, a_1 > 0] \in G\,,$$

$$g_2 = [\boldsymbol{B}_2, a_2 \,:\, \boldsymbol{B}_2 \in \mathcal{GL}(N),\, a_2 > 0] \in G\,,$$

associates the element

$$g_3 = [\boldsymbol{B}_1\boldsymbol{B}_2, a_1 a_2 \,:\, \boldsymbol{B}_1\boldsymbol{B}_2 \in \mathcal{GL}(N),\, a_1 a_2 > 0] \in G\,,$$

called product and denoted as $g_1\, g_2$. It is easy to show that

- *group multiplication* obeys the associative law, i.e.

$$(g_a\, g_b)\, g_c \;=\; g_a\, (g_b\, g_c)\,, \qquad \forall g_a\,, g_b\,, g_c \in G\,;$$

- the element

$$e_1 = [\mathbf{I}_N,\, 1] \in G\,,$$

  is the unique identity, namely $g\, e_1 \;=\; e_1\, g \;=\; g$, $\forall g \in G$;

- $\forall g \in G$ there exists the unique inverse

$$g^{-1} = [\boldsymbol{B}^{-1}, a^{-1} \,:\, \boldsymbol{B}^{-1} \in \mathcal{GL}(N),\, a^{-1} > 0] \in G$$

  such that $g\, g^{-1} = e_1$.

Hence G satisfies the conditions which define a group [16, p. 569].

Appendix D

This appendix is devoted to the derivation of the P_{fa} and P_d for the GLRT exploiting $N = 2$ polarimetric channels. Under either H_0 and H_1, the test statistic in (3.21) is the standard condition number of a matrix following the so-called F distribution [26]. We begin providing its Cumulative Distribution Function (CDF) under both hypotheses in the following

Proposition 2: Under H_0, the CDF of the test statistic (3.21) can be written as

$$\begin{aligned} F_\eta^0(x) = \frac{\beta}{\gamma^{2K}} \Bigg[\int_0^{+\infty} \frac{\lambda^{K-2}}{(1+\lambda/\gamma)^{2K}} \int_\lambda^{x\lambda} \frac{u^K}{(1+u/\gamma)^{2K}} du\, d\lambda - \\ 2\frac{\lambda^{K-1}}{(1+\lambda/\gamma)^{2K}} \int_\lambda^{x\lambda} \frac{u^{K-1}}{(1+u/\gamma)^{2K}} du d\lambda + \\ \frac{\lambda^K}{(1+\lambda/\gamma)^{2K}} \int_\lambda^{x\lambda} \frac{u^{K-2}}{(1+u/\gamma)^{2K}} du d\lambda \Bigg], \end{aligned} \tag{D.1}$$

with $\beta = \frac{(2K-1)}{[(K-2)!]^2} \prod_{r=K}^{2K-2} r^2$. Under H_1,

$$F_\eta^1(x) = \upsilon\beta \int_0^{+\infty} \sum_{\ell=1,k\neq\ell}^{2} \frac{(-1)^{\ell+1}\lambda^{K-2}}{(1+\lambda/\delta_\ell)^{2K-1}} \left[\int_\lambda^{x\lambda} \frac{u^{K-2}(u-\lambda)}{(1+u/\delta_k)^{2K-1}} du \right] d\lambda, \tag{D.2}$$

where $\upsilon = \frac{2K-1}{\delta_1^{K-1}\delta_2^{K-1}(\delta_2-\delta_1)}$, β is the same as in (D.1), and δ_i, $i = 1, 2$, is the i-th eigenvalue of $\mathbf{\Sigma}_X \mathbf{\Sigma}_Y^{-1}$.

Proof. The CDF of the condition number of a matrix-variate whose joint eigenvalues distribution can be written as [38, Formula (6)]

$$f(\lambda) = K \det(\mathbf{\Phi}(\lambda)) \det(\mathbf{\Psi}(\lambda)) \prod_{\ell=1}^{2} \xi(\lambda_\ell)\,, \tag{D.3}$$

can be obtained exploiting [38, Formula (9)]. By inspection of the corresponding joint distribution of λ_1 and λ_2 under H_0, i.e. ([26, Formula (98)] with $\mathbf{\Omega} = \gamma \boldsymbol{I}$)

$$p(\lambda_1, \lambda_2) = \beta/\gamma^{2K} \prod_{i=1}^{2} \frac{\lambda_i^{K-2}}{(1+\lambda_j/\gamma)^{2K}} (\lambda_2 - \lambda_1)^2\,, \tag{D.4}$$

and, respectively, under H_1 (see again [26, Formula (98)])

$$p(\lambda_1, \lambda_2) = \upsilon\beta \prod_{i=1}^{2} \lambda_i^{K-2} \det((1+\lambda_j/\delta_i)^{1-2K})(\lambda_2 - \lambda_1)\,, \tag{D.5}$$

it is immediate to verify that both the above laws are in the form (D.3). Henceforth, under H_0, the CDF takes the form

$$F_0(x) = \beta \int_0^{+\infty} \{\det(\mathbf{F}_1(\lambda)) + \det(\mathbf{F}_2(\lambda))\} d\lambda\,, \tag{D.6}$$

where the entries of the 2×2 matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ are listed below, skipping for simplicity the subscript of λ:

$$(\mathbf{F}_1)_{1,1} = \frac{\lambda^{K-2}}{(1+\lambda/\gamma)^{2K}}, \quad (\mathbf{F}_1)_{1,2} = \frac{\lambda^{K-1}}{(1+\lambda/\gamma)^{2K}},$$

$$(\mathbf{F}_1)_{2,1} = \int_\lambda^{x\lambda} \frac{u^{K-1}}{(1+u/\gamma)^{2K}} du, \quad (\mathbf{F}_1)_{2,2} = \int_\lambda^{x\lambda} \frac{u^{K}}{(1+u/\gamma)^{2K}} du,$$

and

$$(\mathbf{F}_2)_{1,1} = \int_\lambda^{x\lambda} \frac{u^{K-2}}{(1+u/\gamma)^{2K}} du, \quad (\mathbf{F}_2)_{1,2} = \int_\lambda^{x\lambda} \frac{u^{K-1}}{(1+u/\gamma)^{2K}} du,$$

$$(\mathbf{F}_2)_{2,1} = \frac{\lambda^{K-1}}{(1+\lambda/\gamma)^{2K}}, \quad (\mathbf{F}_2)_{2,2} = \frac{\lambda^{K-2}}{(1+\lambda/\gamma)^{2K}}.$$

Under H_1, instead, the CDF can be expressed as

$$F_1(x) = \upsilon\beta \int_0^{+\infty} \{\det(\mathbf{G}_1(\lambda)) + \det(\mathbf{G}_2(\lambda))\} d\lambda \,, \tag{D.7}$$

where the entries of the 2×2 matrix $\mathbf{G}_i$, $i = 1, 2$, are given by

$$(\mathbf{G}_1)_{1,1} = \frac{\lambda^{K-2}}{(1+\lambda/\delta_1)^{2K-1}}, \quad (\mathbf{G}_1)_{1,2} = \frac{\lambda^{K-1}}{(1+\lambda/\delta_1)^{2K-1}},$$

$$(\mathbf{G}_1)_{2,1} = \int_\lambda^{x\lambda} \frac{u^{K-2}}{(1+u/\delta_2)^{2K-1}} du, \quad (\mathbf{G}_1)_{2,2} = \int_\lambda^{x\lambda} \frac{u^{K-1}}{(1+u/\delta_2)^{2K-1}} du,$$

and

$$(\mathbf{G}_2)_{1,1} = \int_\lambda^{x\lambda} \frac{u^{K-2}}{(1+u/\delta_1)^{2K-1}} du, \quad (\mathbf{G}_2)_{1,2} = \int_\lambda^{x\lambda} \frac{u^{K-1}}{(1+u/\delta_1)^{2K-1}} du,$$

$$(\mathbf{G}_2)_{2,1} = \frac{\lambda^{K-2}}{(1+\lambda/\delta_2)^{2K-1}}, \quad (\mathbf{G}_2)_{2,2} = \frac{\lambda^{K-1}}{(1+\lambda/\delta_2)^{2K-1}}.$$

Hence, basic algebra leads to Proposition 2 statement.

As an immediate consequence of Proposition 2, we can now prove the following

Proposition 3: P_{fa} and P_d of the GLRT (3.21) are given by

$$P_{fa} = 1 - \beta \left[\mathcal{J}(0) - 2\mathcal{J}(1) + \mathcal{J}(2)\right] , \tag{D.8}$$

where β is the same as in Proposition 2, T_2 is the detection threshold and, for $\ell = 0, 1, 2$,

$$\mathcal{J}(\ell) = \int_0^{+\infty} \frac{z^{K+\ell-2}}{(1+z)^{2K}} \int_z^{T_2} \frac{y^{K-\ell}}{(1+y)^{2K}} dy dz, \tag{D.9}$$

and by

$$\begin{aligned} P_d =& 1 - (2K-1)\upsilon\beta \int_0^{+\infty} z^{K-2} \left[\int_z^{T_2} \frac{y^{K-2}(y-z)}{(1+y)^{2K-1}} dy\right] \times \\ & \left[\frac{\delta_2^{2K-1}}{(1+z/\omega)^{2K-1}} - \frac{\delta_1^{2K-1}}{(1+z\omega)^{2K-1}}\right] dz, \end{aligned} \tag{D.10}$$

with $\omega = \delta_1/\delta_2$.

Proof. Proposition 3 follows immediately from Proposition 2, expanding the determinant in the argument of (D.4) and (D.5) for P_{fa} and for P_d, respectively. As expected, P_{fa} does not depend on the scaling factor γ and P_d only depends on $\omega = \delta_1/\delta_2$, after replacing υ by its expression as function of δ_i's, $i = 1, 2$.

Bibliography

[1] M. Preiss and N. J. S. Stacy, "Coherent Change Detection: Theoretical Description and Experimental Results", Intelligence, Surveillance and Reconnaissance Division, Defence Science and Technology Organisation, DSTO-TR-1851.

[2] C. Jr. Jakowatz , D. Wahl, P. Eichel, D. Ghiglia, P. Thompson, "Spotlight-mode Synthetic Aperture Radar: A signal processing approach", Kluwer Academic Publishers, BOSTON, MA, 1996.

[3] R. Touzi, A. Lopes, J. Bruniquel, and P. W. Vachon, "Coherence Estimation for SAR Imagery", IEEE Trans. on Geosci. Remote Sens., vol. 37, no. 1, pp. 135-149, Jan. 1999.

[4] I. Stojanovic and L. Novak, "Algorithms Improve Synthetic Aperture Radar Coherent Change Detection Performance", SPIE Newsroom. doi: 10.1117/2.1201307.004889, Jul. 2013.

[5] I. Stojanovic and L. Novak, "Change Detection Experiments Using Gotcha Public Release SAR data", Proc. SPIE 8746, Algorithms for Synthetic Aperture Radar Imagery XX, 87460I, doi:10.1117/12.2020650, Jun. 3, 2013.

[6] L. Novak, "Coherent Change Detection for Multi-Polarization SAR", Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, Oct. 2005, pp. 568-573, doi: 10.1109/ACSSC.2005.1599813.

[7] E. J. M. Rignot and J. J. Van Zyl, "Change Detection Techniques for ERS-1 SAR data", IEEE Trans. on Geosci. Remote Sens., vol. 31, no. 4, pp. 896-906, Jul. 1993.

[8] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A Test Statistic in the Complex Wishart Distribution and Its Application to Change Detection in Polarimetric SAR Data", IEEE Trans. on Geosci. Remote Sens., vol. 41, no. 1, pp. 4-19, Jan. 2003.

[9] L. M. Novak, "Change Detection for Multi-polarization, Multi-pass SAR", SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery XII, Orlando, FL, pp. 234-246, Mar. 2005.

[10] A. Reigber et al., "Very-High-Resolution Airborne Synthetic Aperture Radar Imaging: Signal Processing and Applications", Proceeding of the IEEE, vol. 101, no. 3, pp. 759-783, Mar. 2013.

[11] A. Marino, S. R. Cloude, and J. M. Lopez-Sanchez, "A New Polarimetric Change Detector in Radar Imagery", IEEE Trans. Geosci. Remote Sens., vol. 51, no. 5, pp. 2986-3000, May 2013.

[12] A. A. Nielsen, "Change Detection in Multispectral Bi-Temporal Spatial Data Using Orthogonal Transformations", in Proc. 8th Austral-Asian Sensing Conf., Canberra ACT, Australia, Mar. 25-29, 1996.

[13] A. A. Nielsen, R. Larsen, and H. Skriver, "Change Detection in Bi-Temporal EMISAR data from Kalø, Denmark, by Means of Canonical Correlations Analysis", in Proc. 3rd Int. Airborne Remote Sensing Conf. and Exhibition, Copenhagen, Denmark, Jul. 7-10, 1997.

[14] E. Erten, A. Reigber, L. Ferro-Famil, and O. Hellwich, "A New Coherent Similarity Measure for Temporal Multichannel Scene Characterization", IEEE Trans. on Geosci. Remote Sens., vol. 50, no. 7, pp. 2839-2851, Jul. 2012.

[15] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[16] E. L. Lehmann, *Testing Statistical Hypotheses*, Springer-Verlag, 2nd edition, 1986.

[17] S. Bose and A. O. Steinhardt, "A Maximal Invariant Framework for Adaptive Detection with Structured and Unstructured Covariance Matrices", IEEE Trans. on Signal Processing, Vol. 43, No. 9, pp. 2164-2175, Sept. 1995.

[18] A. De Maio and E. Conte, "Adaptive Detection in Gaussian Interference With Unknown Covariance After Reduction by Invariance", IEEE Trans. on Signal Processing, Vol. 58, no. 6, pp. 2925-2934, Jun. 2010.

[19] A. De Maio, E. Conte and C. Galdi, "CFAR Detection of Multidimensional Signals: an Invariant Approach", IEEE Trans. on Signal Processing, Vol. 58, no. 1, pp. 142-151, Jan. 2003.

[20] U.S. Air Force Sensor Data Management System, "Coherent Change Detection Challenge Problem", https://www.sdms.afrl.af.mil/index.php?collection=ccd_challenge.

[21] S. Scarborough, L. Gorham, M. J. Minardi, U. K. Majumdar, M. G. Judge, L. Moore, L. Novak, S. Jaroszewski, L. Spoldi, and A. Pieramico, "A Challenge Problem for SAR Change Detection and Data Compression", Proc. SPIE 7699, Algorithms for Synthetic Aperture Radar Imagery XVII, 76990U, doi:10.1117/12.855378 , Apr. 18, 2010.

[22] L. M. Novak and M.C. Burl, "Optimal speckle reduction in polarimetric SAR imagery", IEEE Trans. on Aerospace and Electronic Systems, vol.26, no.2, pp.293-305, March 1990.

[23] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the Equivalent Number of Looks in Polarimetric Synthetic Aperture Radar Imagery", IEEE Trans. Geosci. Remote Sens., vol. 47, no. 11, pp. 3795-3809, Nov. 2009.

[24] C. Lopez-Martinez, "Speckle Noise Characterization and Filtering in Polarimetric SAR Data", https://earth.esa.int/documents/10174/669756/ Speckle_Noise_Characterisation.pdf.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Vol. I, Prentice-Hall PTR, 1998.

[26] A. T. James, "Distribution of Matrix Variates and Latent Roots Derived from Normal Samples", Annales of Mathematical Statistics, Vol. 35, pp. 475-501, 1964.

[27] L. L. Scharf, *Statistical Signal Processing. Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1991.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, MA: Cambridge Univ. Press, 1985.

[29] T. S. Ferguson, *Mathematical Statistics*, Academic, New York 1967.

[30] A. Yu. Orlov, "New Solvable Matrix Integrals", International Journal of Modern Physics A, Vol. 19, pp. 276-293, 2004.

[31] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Pt. 1, John Wiley & Sons, 1968.

[32] S. M. Kay and J. R. Gabriel, "An Invariance Property of the Generalized Likelihood Ratio Test", IEEE Signal Processing Letters, Vol. 10, no 12, pp. 352-355, Dec. 2003.

[33] V. Akbari, S. Anfinsen, A. Doulgeris, and T. Eltoft, "The Hotelling-Lawley Trace Statistic for Change Detection in Polarimetric SAR Data under the Complex Wishart Distribution," IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2013, pp. 4162-4165, 21-26 July 2013, Melbourne, Australia

[34] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*, Scitech Publishing, Raleigh NC, 2010.

[35] A. Marino and I. Hajnsek, "A Change Detector Based on an Optimization with Polarimetric SAR Imagery", IEEE Trans. Geosci. Remote Sens., vol. 52, no. 8, pp. 4781-4798, Aug. 2014.

[36] J. Inglada and G. Mercier, "A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and Its Extension to Multiscale Change Analysis", IEEE Trans. Geosci. Remote Sens., vol. 45, no. 5, pp. 1432-1445, May 2007.

[37] S. H. Simon, A. L. Moustakas, and L. Marinelli, "Capacity and Character Expansions: Moment-Generating Function and Other Exact Results for MIMO Correlated Channels", IEEE Trans. on Information Theory, Vol. 52, no. 12, pp. 5336-5351, Dec. 2006.

[38] M. Matthaiou, M. R. Mc Kay, P. J. Smith, and J. A. Nossek, "On the Condition Number Distribution of Complex Wishart Matrices", IEEE Trans. Comm., vol. 58, no. 6, pp. 1705-1717, Jun. 2010.